Applicability of toxicogenomics in molecular epidemiology:

Gene expression profiling as a new biomarker

Danitsja Maria van Leeuwen

ISBN-10: 90-5278-581-3
ISBN-13: 978-90-5278-581-3
Universitaire Pers Maastricht

© Danitsja M. van Leeuwen, Maastricht 2006
Omslagillustratie en lay-out: Danitsja M. van Leeuwen
Omslagrealisatie en druk: Datawyse Boekproducties, Maastricht

nutrim

The studies presented in this thesis were performed at the Nutrition and Toxicology Research Institute Maastricht (NUTRIM), which participates in the Graduate School VLAG (Food Technology, Agrobiotechnology, Nutrition and Health Sciences), accredited by the Royal Netherlands Academy of Arts and Sciences.

Applicability of toxicogenomics in molecular epidemiology:

Gene expression profiling as a new biomarker

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen op
donderdag 7 december 2006 om 14:00 uur

door

Danitsja Maria van Leeuwen

geboren te Delft op 12 mei 1979

Promotor:

Prof. dr. J.C.S. Kleinjans

Copromotor:

Dr. J.H.M. van Delft

Beoordelingscommissie:

Prof. dr. J.P.M. Geraedts (voorzitter)

Prof. dr. W.F. Baeyens (Vrije Universiteit Brussel, België)

Prof. dr. E.C.M. Mariman

Prof. dr. C.P. van Schayck

Prof. dr. E.F.M. Wouters

The research presented in this thesis was conducted as part of and therefore financed by the Flanders Environment and Health Study (Steunpunt Milieu en Gezondheid), currently known as the Flemish Center of Expertise of Environment and Health.

Financial support for the publication of this thesis by the Dr. Ir. van de Laar Stichting is gratefully acknowledged.

Contents

<i>Chapter 1</i>	General Introduction	11
	1.1 Environmental chemical carcinogenesis	
	1.2 Molecular cancer epidemiology	
	1.3 Biomarkers for human biomonitoring	
	1.3.1 Biomarker characteristics	
	1.3.2 Limitations to the current biomarkers	
	1.4 Toxicogenomics	
	1.5 Gene expression profiling technologies	
	1.5.1 Microarray technology	
	1.5.2 Real-time PCR technology	
	1.6 Toxicogenomics in molecular epidemiology	
	1.6.1 Gene expression profiling in human cells in vitro	
	1.6.2 Gene expression profiling in human populations	
	1.7 Aim and outline of the thesis	
<i>Chapter 2</i>	Differential gene expression in human peripheral blood mononuclear cells induced by cigarette smoke and its constituents	31
<i>Chapter 3</i>	Cigarette smoke-induced differential gene expression in blood cells from monozygotic twin pairs	53
<i>Chapter 4</i>	Comparison of cigarette smoke-induced transcriptional responses in human airway epithelium and in blood cells	69
<i>Chapter 5</i>	Genome-wide differential gene expression in children exposed to air pollution in the Czech Republic	89
<i>Chapter 6</i>	Biomonitoring of differential gene expressions in peripheral blood in selected environmentally exposed human populations in Flanders	107
<i>Chapter 7</i>	Summary and General Discussion	129
	Samenvatting	137
	Dankwoord	145
	Curriculum vitae	149
	List of publications and achievements	153

Abbreviations

aa	amino-allyl
AA	aromatic amine
4-ABP	4-aminobiphenyl
AC#	GenBank Accession number
BaP	benzo[a]pyrene
cDNA	copy or complementary deoxyribonucleic acid
CSC	cigarette smoke condensate
Ct	threshold cycle
Cy3	Cyanine-3
Cy5	Cyanine-5
DAVID	Database for Annotation, Visualization and Integrated Discovery
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
EASE	Expression Analysis Systematic Explorer
<i>e.g.</i>	<i>exempli gratia</i> (for example)
FACS	fluorescence activated cell sorting
FLEHS	Flanders Environment and Health Study
GEPAS	Gene Expression Pattern Analysis Suite
H ₂ O ₂	hydrogen peroxide
HPLC	high performance liquid chromatography
IARC	International Agency for Research on Cancer
<i>i.e.</i>	<i>id est</i> (that is)
LOWESS	locally weighted scatterplot smoothing
MN	micronuclei
NADP	nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NHBE	normal human bronchial epithelial cells
NNK	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
PAH	polycyclic aromatic hydrocarbon
PBMC	peripheral blood mononuclear cells
PBS	phosphate buffered saline
PI	propidium iodide
PM	particulate matter
(m)RNA	(messenger) ribonucleic acid
RT	reverse transcription
RT-PCR	real-time polymerase chain reaction (qPCR)
SDS	sodium dodecyl sulphate
se/SE	standard error of the mean
SSC	saline sodium citrate
TIGR	the Institute for Genomics Research
TPM	total particulate matter
v/v	volume/volume

Chapter 1

General Introduction

1.1 Environmental chemical carcinogenesis

A good health demands a healthy environment. Over the last decades it has been of major interest to understand how and to what extent humans influence, or pollute, their environment and how in turn, the environment influences human health. The development of civilization and the emergence of industries over the preceding centuries have greatly contributed to the increased presence of complex mixtures of hazardous toxic and carcinogenic chemicals in the human environment, many of which are suspected or have been proven to cause cancer. Human exposure to environmental carcinogens may result from several sources, related to the general environment, occupational settings, and dietary or life style habits (*e.g.* cigarette smoking). Today, cancer is one of the leading causes of death worldwide, with annually increasing numbers of patients. A 2005 publication by the International Association on Cancer Research (IARC, Lyon, France) [1] reports nearly 2.9 million incident cases of cancer diagnosis as well as 1.7 million cancer deaths in Europe in 2004. Of all cancers, lung cancer is the most common incident form, as well as, by far the most common cause of death of cancer.

The process of cancer development as a result of exposure to environmental chemicals is known as environmental chemical carcinogenesis and comprises a complex, multistage process that may take years or even decades to become manifest. The carcinogenic process is shown schematically in Figure 1 and elucidates the transformation of a normal 'healthy' cell into a malignant tumor [2]. Roughly, cells pass four stages before clinical appearance of a tumor: namely initiation, promotion, progression and transformation.

The carcinogenic process encompasses a series of interactions between exogenous (environmental) factors and endogenous processes. Each of these interactions may be influenced by the genetic background of the exposed individual [3]. There is strong evidence that supports the substantial contribution of environmental factors to the overall outcome of the steps in the carcinogenic process, indicating environmental exposure to account for more than 50% of the cancer risk [4-7]. As a considerable part of cancers has an environmental component, in theory that part of the incident cancer cases should thus be preventable [8]. It is therefore of utmost importance to further examine chemical carcinogenesis, to gain better and more understanding of this process. This will aid in better, more efficient definition of the risks posed by environmental chemicals which will very likely result in better prevention of the disease.

Adapted from Harris [2]

Figure 1. Simplified overview of the multistage process of carcinogenesis.

1.2 Molecular cancer epidemiology

Classical epidemiology has been the traditional approach to study associations between exposure to hazardous substances and the development of diseases such as cancer, using disease incidence and mortality as endpoints. For several reasons, however, conventional epidemiological approaches for studying human cancer risks from environmental exposure are not practical: the long development period of the disease following exposure to carcinogens, the low frequency / incidence of the disease and the numbers of study subjects therefore required, are of disadvantage in this respect [9].

In 1982, Perera and Weinstein [10] proposed for the first time the concept of molecular cancer epidemiology [11-13], a research approach developed with the goal to study cancer causation in humans at a mechanistic level. A framework was provided for the incorporation of molecular and biological endpoints into classical epidemiology, with the pivotal role to elucidate the biochemical and molecular aspects of cancer etiology. To this end, individuals from populations at elevated cancer risk are studied for key pre-carcinogenic events in their body which in general occur at the earliest stages of carcinogenesis, *i.e.* long before a visible or diagnostic health effect appears. Figure 2 shows the sequence of molecular and biological events occurring in the human body, preceding the clinical manifestation of cancer, which are currently identified to be indicative for cancer risk and are in use for human biomonitoring [14]. These key events are referred to as biomarkers [15-17] and are defined in more detail as 'xenobiotically induced variations in biological, cellular or molecular components, processes, structures or functions that are measurable in a biological system or matrix' [18].

The possibilities of applying these biomarkers to substitute for the classical endpoints of cancer incidence and mortality in epidemiological studies, represent the major feature of molecular cancer epidemiology [19,20]. In molecular cancer epidemiology, improved detection and prevention have been considered as long-standing principal goals [21,22]. Molecular cancer epidemiology overcomes the relative disadvantages of classical cancer epidemiology by the use of biomarkers which provide the possibility of an assessment of the cancer risk at very early stages of the carcinogenic process, eliminating the 'need' to await clinical manifestation of a malignancy. The use of biological markers comprises and reflects individual differences with regard to the fate of the exposure chemical inside the human body and therefore provides a more accurate measure of the exposure and its effects. The next paragraph will further elucidate different types of biomarkers.

Figure 2. The conceptual basis of biomarkers for use in molecular cancer epidemiology. Individual genetic susceptibility factors can influence all the stages between exposure and onset of disease.

1.3 Biomarkers for human biomonitoring

In research on environmental toxicology, there is considerable interest in the use of biomarkers as indicators of early changes that possibly lead to clinical disease later [23,24]. This has been reflected in a number of reviews that have highlighted the important role of biomarkers in contemporary toxicology [24-28]. Preferably, in molecular epidemiology biomarkers are applied that reflect changes at the earliest stages of the carcinogenic process (Figures 1 and 2), and biomarker analysis occurs among (apparently) healthy humans and human populations. This paragraph describes classes of biomarkers in chronological order in the carcinogenic process.

Following the paradigm as displayed in Figure 2, environmental monitoring is the first monitoring step in the process of chemical carcinogenesis and refers to the assessment of exposure. This is achieved by measuring (monitoring) the presence and concentration of chemicals in environmental media *e.g.* ambient air, and is widely practiced. However, by presenting information only on the occurrence of substances in an environmental matrix, environmental monitoring merely provides an estimate of the amount that might enter the human body and may cause adverse effects. It does not reflect internal exposure, as it does not take into account absorption, distribution and metabolism of chemicals, which can differ drastically between chemicals and between individuals.

Biological monitoring, as compared to environmental monitoring, is generally more informative for cancer risk assessment as it comprises measurement of the internal dose (biological monitoring), the biologically effective dose as well as the early biological effect (biological effect monitoring, Figure 2). These markers take into account (inter-individual differences in) absorption, metabolism and distribution of the carcinogen and are therefore more relevant to assessment of risks [29]. Internal dose markers are frequently referred to as biomarkers of internal exposure. These markers are measured in body fluids or excreta. An example of a widely-used biomarker of internal dose for cigarette smoke exposure is the assessment of the cotinine concentration in urine or blood plasma. Cotinine is the primary metabolite of nicotine and is referred to as a better biomarker of exposure to cigarette smoke than its parent compound nicotine as it has a longer half-life in the body [30]. Furthermore, cotinine is a biomarker with a very high specificity as it is unique for tobacco smoke exposure.

Because of practical and ethical reasons in studying humans, biological effect monitoring relies on the use of surrogate target cells, *i.e.* easily accessible tissue that can be obtained in a relatively non-invasive manner. Surrogate target cells are cells that are investigated as surrogate for the cells of target tissues and organs for carcinogenesis, which mostly are not accessible for biomonitoring studies. White blood cells are widely used for the measurement of biomarkers of effect [31-33]. In traditional human biomonitoring studies, white blood cells or lymphocytes are preferred for the study of genotoxic effects [31].

Among the biological effect markers, biomarkers of the biologically effective dose and biomarkers of early biological effects can be distinguished, most of which are in general measured in blood. Examples of effective dose markers are DNA and protein adducts. DNA adducts are formed through the covalent binding of chemicals to specific

sites in DNA, thus forming a readily measurable biomarker by *e.g.* ^{32}P -postlabeling [34,35]. DNA adduct analysis has become very common in molecular epidemiology as DNA adducts are likely to lead to mutations through erroneous replication [36]. Thus, among the biomarkers of effective dose, DNA adducts are clearly relevant with regard to the disease process. Effective dose markers can in general be reversible, among others because they can be repaired during the cell's life. Adducted sites in the DNA, for instance, can be repaired by the cell's DNA repair machinery before the adduct results in a (critical) mutation. Furthermore, DNA adducts have been validated as indicative of cancer risk, as significantly elevated levels of DNA adducts have been found in blood cells of cigarette smokers who developed lung cancer [37].

Early biological effect markers represent in general an irreversible event, *i.e.* one that cannot be repaired. Biomarkers of early biological effect comprise *e.g.* mutations in the HPRT gene and cytogenetic damage, such as chromosome aberrations (CA) and the formation of micronuclei (MN). Cytogenetic biomarkers are the most frequently used surrogate endpoints for cancer risk in human population studies [38]. In 2002, Bonassi and Au have reviewed CA as the most extensively studied and applied biomarker of early biological effects in population studies [19,39]. The relevance of these cytogenetic damages to the development of cancer is clearly indicated by the presence of chromosome deletions and translocations in the majority of cancer cells [40]. Recently, CA and MN have been causally associated with the development of cancer [38,41], whereas another early biological effect marker (high-frequency sister chromatid exchange cells) was not associated with the risk of cancer [42]. Nowadays, the MN assay is practiced more frequently than the CA assay, as it is less labor-intensive [27].

Genetically determined variations (polymorphisms) in genes coding for enzymes involved in carcinogenesis specific processes are very likely to affect individual risk and inter-individual differences in biomarker levels [22]. Factually, genetic variation in susceptibility for the effects of environmental exposure may affect each of the steps on the causal pathway from exposure to disease (Figure 2) and is therefore an important aspect in human biomonitoring. Since many human carcinogens require metabolic activation to generate the ultimately reactive carcinogenic metabolites, polymorphisms in genes encoding metabolic enzymes are of importance for the individual's susceptibility for all biomarkers. This is also of relevance *e.g.* for polymorphisms in genes coding for DNA repair enzymes, which however may affect only the biomarkers for biologically effective dose and the biomarkers for early biological effects. It is apparent that biomarkers of exposure, effect and susceptibility are closely interconnected. Conceptually, the biomarkers are approached separately (Figure 2) but in reality they form a continuum [43].

Summarizing, biomarkers are used to identify and quantify chemical exposures and effects predictive of cancer risk. In this way they aid in cancer risk assessment of humans exposed to environmental carcinogens.

1.3.1 Biomarker characteristics

The ideal biomarker is characterized by (high) sensitivity and specificity, and should also be well-validated. However, it is beyond reality for a biomarker to display 100% sensitivity and 100% specificity. This section defines the concepts of sensitivity, specificity and validity/validation of biomarkers.

Sensitivity of a biomarker is defined as to how significant the biomarker discriminates exposed from non-exposed populations or individuals, or the reliability with which the biomarker classifies a true positive or negative result. A sensitive biomarker is also considered to detect small differences, preferably also in a small population. Biomarkers for the same precarcinogenic effect show different sensitivities. For example, CA are weakly induced by cigarette smoke, which alternatively is a strong inducer of HPRT and lung cancer-related mutations. Vice versa, ethylene oxide is a strong inducer of CA, but induces mutations only weakly [44].

Specificity of a biomarker refers to how unique the measured marker is to either the specific exposure or the specific effect, or the capability of the marker to discriminate true effects from false positive and negative effects. As described previously, measuring cotinine concentration in urine or plasma is a highly specific biomarker of internal dose for cigarette smoke exposure as it is unique to this exposure. Cytogenetic markers, biomarkers of early effect like chromosome aberrations, generally have poor specificity as they do not provide information on the exposure chemical, but high specificity for the endpoint (cancer) [24]. However, they are associated with increased cancer risk irrespective of the carcinogenic exposure. The biomarkers at the far left end in Figure 2, the environmental and biological markers possess a high specificity for the exposure (the chemicals) but a low(er) specificity for the health effect, cancer. Vice versa, the early biological effect markers have a higher specificity for the health effects. The markers of biologically effective dose are in between; *e.g.* DNA adducts reflect the (type of) exposure as well as indicate a health effect with regard to the possible mutations that may result from the adducted sites.

Validity of a biomarker is often defined as the presence of proof that the measured event reliably predicts cancer risk. However, criteria as to what it takes to properly validate a biomarker as a surrogate endpoint are not standardized [20]. Often the development and validation of a biomarker follows a tiered approach that starts with experimental *in vitro* research followed by transitional population studies to validate the biomarker *in vivo* [20,22]. At the start of the present research the only biomarkers that were validated were DNA adducts and CA [45], which indicates the need for the development of new biomarkers to benefit molecular epidemiology.

1.3.2 Limitations to the current biomarkers

Thus, the biomarker approach as is currently in use, clearly strengthens the field of molecular cancer epidemiology. However, a number of (relative) limitations regarding the existing biomarkers can be indicated. In 2000, Perera and Weinstein concluded that many of the biomarkers used in molecular epidemiology needed further validation, meaning that for many the relevance for the chemical carcinogenic process was not proven, but mostly hypothetical.

An overall limitation of biological effect monitoring in molecular epidemiology is the investigation of effects in surrogate target cells rather than cells or tissues that are targets chemical carcinogenesis, such as lung and liver. Molecular epidemiology relies on the use of these cells because of practical and ethical reasons. However, surrogate target cells provide a reflection of the events that occur in the target tissue. A biomarker is more relevant for disease risk when measured in target tissue, but this is almost al-

ways impossible because of the invasive procedures to obtain tissues from *e.g.* lung or liver from health humans. A few papers have addressed the comparison of outcomes of biomarker measurements in surrogate and target tissue. Factually, DNA adduct levels are the only biomarker that has been found to display similar patterns in target and surrogate target cells as found in experimental and human studies [46-48].

The general limitations that were ascribed to the biomarkers at the start of the current project were as follows. First, many assays are labor-intensive and therefore hardly applicable for screening or monitoring populations, which generally comprise a large number of individuals. Second, for the measurement of each intermediate endpoint of environmental carcinogenesis, a separate biomarker assay needs to be applied. In addition, the current biomarkers provide mechanistic information on only one event occurring in the cell during or after exposure. It is very likely that more and other alterations take place which are not detected by means of the ongoing approach. Furthermore, there is no early biomarker that provides mechanistically relevant information on a generic level, *i.e.* on more than one effect (event or mechanism) at a time.

The specificity of certain biomarkers remains to be elucidated. In general there are few current biomarkers that are specific for one chemical or exposure. In addition, the biological significance, *i.e.* representation of an event within the continuum from exposure to disease, of some biomarkers is (yet) unclear. For example, the enthusiasm about the sister chromatid exchange (SCE) assay, generally being more sensitive than the CA assay, has decreased over the past years as the biological mechanism and thus relevance, of SCE formation is elusive [19]. Sensitivity is also a desired characteristic that is subject to improvement, especially in the case of low dose exposures to environmental chemicals, which are most common to the general human population.

Taking into account these disadvantages, it can be argued that molecular cancer epidemiology will benefit from (the development of) new biomarkers that are hypothesized to overcome the relative limitations ascribed to the current markers. These new biomarkers would ideally provide more information, on a more generic and more mechanistic level than the current biomarkers, thereby increasing specificity of environmental cancer risk assessment. They would give insight into 'all' events that occur in the cell or tissue due to chemical exposure, and therefore might be informative on multiple health effects simultaneously, thereby increasing specificity. New biomarkers that also tackle the disadvantageous need to use different methods for each biomarker to be measured would also be of high value. Finally, new biomarkers should be able to detect and measure effects at low exposure levels. Newly emerging technologies, such as toxicogenomics, may provide a promising possibility for the development of new, improved biomarkers, comprising the profiling of changes in the expression of multiple genes simultaneously in consequence to toxicant exposure. The next section will further elaborate on this subject.

1.4 Toxicogenomics

In the early nineties, the field of toxicology recognized that increased understanding had to be gained concerning the mechanism of action of chemicals, in order to improve the efficiency of its tasks: hazard identification and mechanism-based risk assessment [49]. In this aspect, toxicology, and consecutively also molecular cancer epidemiology, is likely

to benefit from the development of the new field of toxicogenomics [49-54]. Toxicogenomics represents an integration of new genomic technologies as well as bioinformatics into conventional toxicology, in an effort to gain knowledge on the role of gene-environment interactions in health effects caused by chemical exposure [55]. It is a rapidly developing discipline that is promising to better understand the effects of chemicals in biological systems at the genome level [49]. The promises and potential of the field of toxicogenomics and its methodologies have been described extensively in an array of papers [49,56-62]. In January 2003, the first and currently only journal specifically dedicated to toxicogenomics, was established as a quarterly edition of *Environmental Health Perspectives* [63], thereby indicating the importance dedicated to this new research area.

The profile of expressed genes determines a cell's or an organism's phenotype. An important assumption in toxicogenomics is that almost without exception, toxicologically relevant effects are accompanied by alterations in gene expression patterns [64,65]. Also, major diseases like cancer are known to be accompanied by altered expression of genes in disease-associated biological pathways, such as metabolism, cell proliferation and apoptosis. Monitoring changes in the expression of (multiple) genes resulting from toxicant or carcinogenic exposure, can provide much valuable information on (the mechanisms underlying) the cellular response to exposure on a broad scale. Several techniques are available for measuring gene expression changes, including Northern blotting, reverse transcription polymerase chain reaction (RT-PCR), differential display of gene expression (DDGE) and serial analysis of gene expression (SAGE) [66-69]. Nowadays, mostly microarrays are used to study simultaneously large parts of the transcriptome, which represents the full complement of messenger RNAs from active genes in a particular tissue, and how that transcriptome responds to xenobiotic stimuli. In 1995, Schena *et al.* [70] were the first to report this technique. The application and characteristics of microarray technology are elucidated in the next paragraph.

1.5 Gene expression profiling technologies

1.5.1 Microarray technology

In general, a microarray is defined as an array of DNA sequences corresponding to genes of interest, immobilized onto a solid surface [71]. Since the earliest days of toxicogenomics, microarrays of different formats have been fabricated commercially or 'in-house', pre-designed or custom-made, from small-numbered dedicated arrays to slides covering the whole genome.

There are two basic types of microarrays, cDNA microarrays and oligonucleotide arrays, which both are used for gene expression profiling but differ in methodology [49,70,72]. cDNA microarrays represent harbor gene sequences in the form of ~200-500 kb nucleotide cDNA fragments that are spotted or printed onto specially coated glass slides. Phase-1 Human Tox 600 cDNA microarrays (Phase-1 Molecular Inc. Santa Fe, New Mexico, USA) as applied in the present project are an example of a commercially available dedicated platform for the simultaneous analysis of 600 toxicology relevant genes. Oligonucleotide microarrays are generated by means of specific chemical synthesis, *e.g.* photolithography or (non-contact) in situ synthesis, to generate the gene-specific sequences, generally consisting of ~20-80 base oligonucleotides or by spotting

of these sequences onto a solid surface. An example of an in-situ synthesized oligonucleotide microarray platform is provided by Agilent (Palo Alto, CA, USA). Their arrays contain 60-mer oligonucleotide sequences of more than 22,000 genes and were also applied in the present project. In general, two concomitant, differentially labeled, samples are competitively hybridized onto the above described types of microarrays.

Microarray technology relies on the use of high quality RNA from any source and an accompanying control; samples originating from either healthy and diseased or exposed and non-exposed subjects. In general, the RNA is reverse transcribed, and for some platforms *in vitro* transcribed for amplification, with either direct or indirect incorporation of a fluorescent label (Figure 3). Competitive hybridization of equal amounts of test and control cDNA or cRNA (probes), each labeled with a different fluorophore, provides signals when probes are bound to the complementary sequences (targets) on the microarray. By means of confocal multiple laser scanning, images are generated which serve to quantify signal intensities based on the number of pixels read in each spot that represents a gene sequence. Besides the types of microarray described which require two differentially labeled samples for hybridization, the Affymetrix platform exists that provides GeneChips which require single sample hybridizations.

1.5.2 Real-time PCR technology

An alternative assay currently frequently used for the analysis of gene expression is real-time PCR. This technique is often applied to confirm microarray data. It is used for (relative) real-time quantification of mRNA transcripts for a defined, limited, number of genes, by means of PCR amplification. In contrast to microarray technology, that is ideal to provide expression data for large numbers of genes in a small number of samples, PCR-based measurements are applied for high-throughput purposes, *i.e.* for expression analysis of a small number of genes in a large number of samples. Like microarray technology, real-time PCR relies on high quality RNA which is reverse transcribed into cDNA. Transcript abundance is then measured by means of amplification of the cDNA combined with measurement of the product formed. Measurement of product can be performed in a multiple ways, such as by staining with SYBR Green, dye that only fluoresces when bound to double-strand DNA. After a set number of amplification rounds, relative transcript abundance values are calculated versus a pooled or control sample.

Figure 3. An overview of the dual-color labeling microarray technology for gene expression profiling using cDNA microarrays

1.6 Toxicogenomics in molecular epidemiology

An advantage of toxicogenomics with the use of large-scale gene expression profiling is that it can be applied to the development of new biomarkers for use in human biomonitoring studies. Microarray technology provides the possibility to identify patterns of gene expression, confirming chemical exposure and possibly providing a chemical-specific profile or fingerprint. In addition, microarray-derived gene expression data provide a wealth of information at the mechanistic level. By generating an overview on the genomic response to environmental carcinogens, an improved understanding of the molecular basis of the detrimental effects caused by the carcinogens can be obtained. Several publications have described the application of toxicogenomics with regard to the development of biomarkers for human biomonitoring [43,55,56,73]. The next section provides an overview of selected currently available reports on gene expression profiling in human biological samples after (environmental) exposure.

1.6.1 Gene expression profiling in human cells *in vitro*

The next section will focus solely on toxicogenomic analyses in human blood cells (as surrogate target cells) and in primary human lung tissue or lung cells (as target cells), since these cell types are most meaningful for research on biomarkers for human molecular cancer epidemiology.

SURROGATE TARGET CELLS

As mentioned earlier, blood cells, especially peripheral blood lymphocytes (PBL) and peripheral blood mononuclear cells (PBMC), are the surrogate target cells of choice for molecular cancer epidemiology. Ryder *et al.* [74] have exposed human PBMC to cigarette smoke *in vitro* with the aim to investigate the effects of acute exposure on gene expression and the implications of this for periodontal disease. As significant modulation of genes was identified in smoke-exposed relative to non-exposed samples the results of this study have elucidated molecular mechanisms that determine the host response to cigarette smoke. Amundson *et al.* [75] investigated changed mRNA levels in human PBL as potential biomarkers for radiation induced effects. Several genes were identified with a dose-dependent increased expression after *ex vivo* radiation. Their results suggested that environmental radiation exposure could be estimated by measuring the relative gene expression levels in PBL.

TARGET CELLS

The major entry route for many carcinogenic chemicals is through inhalation. As a result, the lung is the first organ to encounter a substantial amount and variety of exposure chemicals. For this reason, human lung cells are the target cells that are considered most relevant to environmental exposure and that will be discussed in this paragraph.

Hellerman *et al.* [76] have reported altered expressions of selected genes in normal human bronchial epithelial (NHBE) cells and A549 (type II lung cell carcinoma) cells after *in vitro* exposure to a smoke condensate of Kentucky 1R3F reference cigarettes. Increased expression of inflammatory mediators was observed. Fields *et al.* [77] have

addressed transcriptional modulation of selected genes involved in DNA damage, oxidative stress, inflammatory response and cell cycle regulation, in NHBE cells caused by different smoke condensates from tobacco-burning and tobacco-heating cigarettes. Significant induction of gene expression after exposure to tobacco-burning cigarettes (Kentucky reference) as compared to the primarily heating cigarettes was reported, however effects after only a short exposure time were examined. Jorgensen *et al.* exposed NHBE cells from a single donor to two different types of cigarette smoke condensate and observed that large numbers of genes were affected, including genes related to carcinogenesis [78].

In theory, the most relevant cell type for *in vitro* research on gene expression profiles in target tissue for carcinogenic effects of environmental exposure concerns fresh or primary cells. However, in practice these are often difficult to obtain and therefore less available. Evidently, investigation of gene expression at target tissue level is most relevant to analysis of the risks and effects of carcinogenic exposure, however for biomarker research the focus is on easily obtainable blood cells. At present, few reports exist on comparisons of gene expression profiles in human surrogate and target tissues. Sullivan *et al.* have recently evaluated comparability of blood and brain gene expression and concluded that blood gene expression neither perfectly resembled nor completely differed from gene expression in brain [79].

1.6.2 Gene expression profiling in human populations

A limited number of studies have been reported up till now, which focused on identification of altered gene expression profiles in a human population exposed to carcinogens or toxicants. These studies have been conducted on populations with well-defined occupational, environmental or life-style related exposures. A transcriptomic analysis of normal human lung tissue was performed by Spira *et al.* [80,81]. Altered gene expression has been observed in the transcriptome of normal airway, in smoking subjects and in former smokers compared to non-smokers, giving insight into the cigarette smoke-induced regulation of gene expression *in vivo*. Forrest *et al.* conducted a study in a population of benzene-exposed shoe-factory workers in China. A matched analysis of 6 exposed-control pairs was performed, and differentially expressed genes in PBL were identified [82]. It was concluded that the altered expressions of a selected set of genes can be potential biomarkers for benzene exposure and that microarray analysis can be a valuable tool for the discovery of mechanistic relevant genes. Thus it seems that these gene expressions can be biomarkers of exposure and early effect in molecular epidemiology studies of populations exposed to benzene. In addition, Wang *et al.* investigated differential gene expression in whole-blood from humans occupationally exposed to welding fumes, *i.e.* inhaled particles [83]. It was demonstrated that with a paired sampling study design of pre- and post-exposure measurements, small changes in gene expression can be measured in a human population, and therefore that the technique can be applied to evaluate the host response to other forms of environmental exposures. Furthermore, Lampe *et al.* showed in an observational study, that it is possible to distinguish smokers from non-smokers on the basis of the gene expression in peripheral leukocytes [84]. Exposure to tobacco smoke was shown to be associated with a biologically relevant gene expression signature as the selected genes correlated with plasma cotinine levels,

a marker tobacco smoke exposure. Wu *et al.* conducted an exploratory study using home-made dedicated cDNA microarrays to identify genes differentially regulated in PBL from arsenic-exposed but apparently healthy humans [85]. This study reports elevated expression levels of genes involved in inflammation and immune response. This again demonstrated the possibility to identify exposure-induced gene expression modulations using microarray technology. In summary, gene expression profiling by microarray technology appears a promising tool for molecular epidemiological investigations of human population exposed to chemical carcinogens.

1.7 Aim and outline of the thesis

In addition to the established and widely used biomarkers that are available to monitor humans for the effects of hazardous environmental exposure that might lead to the development of cancer, molecular epidemiology will greatly benefit from the development of new biomarkers that would give more generic, biologically relevant and mechanistic information on the events that occur in the cell at the time of or as a result from the exposure. It is known that the detrimental response of a cell or an organism to carcinogenic substances, like components present in cigarette smoke or air pollution, is partly the result of alterations in the expression of genes. Since gene expression is a highly regulated process, monitoring the changes of large numbers of genes may give an impression of all the events at the genome level, thereby offering biologically meaningful information on multiple pre-carcinogenic events simultaneously. Furthermore, gene expression is likely to be modulated in response to low exposure levels, suggesting gene expression profiling to be a more sensitive measure than current biomarkers used for the same purpose of environmental cancer risk assessment.

The aim of the research presented in this thesis was to develop and confirm gene expression profiles as potentially new biomarkers of early biological response to exposure to chemical carcinogens thereby hypothesizing that (environmental) exposure to carcinogenic substances will cause alterations at the transcriptome level. Concomitantly, it was the goal to apply the retrieved gene expression profile in a biomonitoring study in surrogate target cells of in human populations with diverse exposures to mixtures of environmental carcinogens.

To embed 'gene expression profiling' in the continuum of biomarkers for use in molecular cancer epidemiology (Figure 2), it is suggested to occupy a new box named 'early biological response', between the biomarkers of biological effective dose and the biomarkers of early biological effect. Altered gene expression is considered a more mechanistically relevant event than the formation of DNA adducts, as this reflects the presence of an agent with DNA binding capacity and the amount of which has bound to the DNA. Since gene expression is dependent on transcription activation, by activation of receptors or more directly, it is considered a 'later' event than the formation of DNA adducts. However altered gene expression is not as irreversible as early effect markers like micronuclei, and should therefore better be named a response than an effect.

The approach that was followed in the present thesis consisted of an *in vitro* and an *in vivo* part. Cigarette smoke was used as a model for carcinogenic substances, which allowed to perform experimental *in vitro* studies and subsequent comparison of these studies to observations in human smokers *in vivo*. Subsequently, two case studies

were conducted in human populations, *i.e.* children and adults, applying gene expression markers to study the effects of ambient air pollution, including carcinogenic compounds, at the transcriptome level.

The outline of this thesis is as follows:

- An initial, exploratory investigation was conducted *in vitro* in human peripheral blood mononuclear cells, and described in **Chapter 2** [86]. Objectives were 1) to identify genes that represented a general response to genotoxic stress by investigating in the surrogate target tissue of choice, which genes respond to the agents tested and 2) to identify the most sensitively responding agent-specific genes that might mark low exposure levels.
- A human population study was performed with the purpose to compare *in vivo* data with those observed *in vitro* and described in **Chapter 3** [87]. Objectives were 1) to analyze blood gene expression differences within smoking-discordant monozygotic twins, in order to identify genes that are deregulated by tobacco smoke in surrogate tissue, and 2) to phenotypically anchor the gene expressions to a biomarker for biological effective dose (conform Figure 2) of exposure to cigarette smoke, *i.e.* DNA-adduct levels.
- An additional *in vitro* study is presented in **Chapter 4**, which focused on identifying cigarette smoke-induced differential gene expression in target tissue (normal primary human bronchial epithelial cells), with the aim to compare the findings in target tissue with those from the surrogate target tissue study of Chapters 2 and 3.
- **Chapter 5** describes the first of two case studies, concentrating on children exposed to ambient air pollution [88]. The objectives were 1) to investigate the potential of a profile of gene expressions to discriminate exposed from non-exposed human populations, and 2) to phenotypically anchor the gene expressions to a biomarker of early biological effect (conform Figure 2), *i.e.* micronuclei frequencies.
- A second case study, described in **Chapter 6**, applied the gene expression profile retrieved from the previous studies in a large adult biomonitoring study. Objectives were 1) to examine the potential of the selected gene expressions to segregate populations with (small) differences in environmental exposure, and 2) to associate the behavior of the expression of these genes with biomarkers of biological effective dose and early biological effect like DNA strand breaks and micronuclei.
- Finally, a summary and general discussion is presented in **Chapter 7**.

References

1. Boyle, P. and Ferlay, J. (2005). Cancer incidence and mortality in Europe, 2004. *Ann Oncol* 16, 481-488.
2. Harris, C.C. (1989). Interindividual variation among humans in carcinogen metabolism, DNA adduct formation and DNA repair. *Carcinogenesis* 10, 1563-1566.
3. Kyrtopoulos, S.A. (2005). Biomarkers in environmental carcinogenesis research: Striving for a new momentum. *Toxicol Lett*
4. Le Marchand, L. (2005). The predominance of the environment over genes in cancer causation: implications for genetic epidemiology. *Cancer Epidemiol Biomarkers Prev* 14, 1037-1039.
5. Luch, A. (2005). Nature and nurture - lessons from chemical carcinogenesis. *Nat Rev Cancer* 5, 113-125.
6. Czene, K., Lichtenstein, P. and Hemminki, K. (2002). Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer* 99, 260-266.
7. Higgingson, J. (1993). Environmental carcinogenesis. *Cancer* 72, 971-977.
8. Weinstein, I.B., Santanella, R.M. and Perera, F.P. The molecular biology and molecular epidemiology of cancer., in: Greenwald, P. (Ed.), *The Science and Practice of Cancer Prevention and Control*, Marcel-Dekker, New York, NY, 1993.
9. Albertini, R.J., Nicklas, J.A. and O'Neill, J.P. (1996). Future research directions for evaluating human genetic and cancer risk from environmental exposures. *Environ Health Perspect* 104 Suppl 3, 503-510.
10. Perera, F.P., Poirier, M.C., Yuspa, S.H., Nakayama, J., Jaretzki, A., Curnen, M.M., Knowles, D.M. and Weinstein, I.B. (1982). A pilot project in molecular cancer epidemiology: determination of benzo[a]pyrene--DNA adducts in animal and human tissues by immunoassays. *Carcinogenesis* 3, 1405-1410.
11. Shields, P.G. and Harris, C.C. (1991). Molecular epidemiology and the genetics of environmental cancer. *Jama* 266, 681-687.
12. Perera, F.P. (1987). Molecular cancer epidemiology: a new tool in cancer prevention. *J Natl Cancer Inst* 78, 887-898.
13. Ambrosone, C.B. and Kadlubar, F.F. (1997). Toward an integrated approach to molecular epidemiology. *Am J Epidemiol* 146, 912-918.
14. Wogan, G.N. (1992). Molecular epidemiology in cancer risk assessment and prevention: recent progress and avenues for future research. *Environ Health Perspect* 98, 167-178.
15. Albertini, R.J. (1999). Biomarker responses in human populations: towards a worldwide map. *Mutat Res* 428, 217-226.
16. Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N.E. IARC Scientific Publications No. 142, International Agency for Research on Cancer, Lyon, 1997.
17. Vanden Heuvel, J.P., Davis, J.W. (1999). Molecular approaches to identify exposure and risk to specific environmental pollutants. *Biomarkers* 4, 93-105.
18. NRC (National Research Council). *Biologic Markers in Reproductive Toxicology*, National Academy Press, Washington DC, 1989.
19. Bonassi, S. and Au, W.W. (2002). Biomarkers in molecular epidemiology studies for health risk prediction. *Mutat. Res* 511, 73-86.
20. Bonassi, S., Neri, M. and Puntoni, R. (2001). Validation of biomarkers as early predictors of disease. *Mutat Res* 480-481, 349-358.
21. Albertini, R.J. (2001). Developing sustainable studies on environmental health. *Mutat Res* 480-481, 317-331.
22. Bartsch, H. (2000). Studies on biomarkers in cancer etiology and prevention: a summary and challenge of 20 years of interdisciplinary research. *Mutat Res* 462, 255-279.
23. Mutti, A. (1995). Use of intermediate end-points to prevent long-term outcomes. *Toxicol Lett* 77, 121-125.
24. Watson, W.P. and Mutti, A. (2004). Role of biomarkers in monitoring exposures to chemicals: present position, future prospects. *Biomarkers* 9, 211-242.
25. Timbrell, J.A. (1998). Biomarkers in toxicology. *Toxicology* 129, 1-12.
26. Mutti, A. (1999). Biological monitoring in occupational and environmental toxicology. *Toxicol Lett* 108, 77-89.
27. Fenech, M. (2002). Biomarkers of genetic damage for cancer epidemiology. *Toxicology* 181-182, 411-416.
28. van Delft, J.H., Baan, R.A. and Roza, L. (1998). Biological effect markers for exposure to carcinogenic compound and their relevance for risk assessment. *Crit. Rev. Toxicol* 28, 477-510.
29. Desikan, R., S, A.H.-M., Hancock, J.T. and Neill, S.J. (2001). Regulation of the Arabidopsis transcriptome by oxidative stress. *Plant Physiol* 127, 159-172.
30. IARC IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans. Volume 38. Tobacco smoking, IARC, Switzerland, 1986.

31. Faust, F., Kassie, F., Knasmüller, S., Kevekordes, S. and Mersch-Sundermann, V. (2004). Use of primary blood cells for the assessment of exposure to occupational genotoxicants in human biomonitoring studies. *Toxicology* 198, 341-350.
32. Rockett, J.C., Burczynski, M.E., Fornace, A.J., Herrmann, P.C., Krawetz, S.A. and Dix, D.J. (2004). Surrogate tissue analysis: monitoring toxicant exposure and health status of inaccessible tissues through the analysis of accessible tissues and cells. *Toxicol. Appl. Pharmacol* 194, 189-199.
33. Rockett, J.C. (2002). Surrogate tissue analysis for monitoring the degree and impact of exposures in agricultural workers. *AgBiotechNet* 4, 1-7.
34. Kriek, E., Rojas, M., Alexandrov, K. and Bartsch, H. (1998). Polycyclic aromatic hydrocarbon-DNA adducts in humans: relevance as biomarkers for exposure and cancer risk. *Mutat Res* 400, 215-231.
35. Miller, E.C. and Miller, J.A. (1981). Mechanisms of chemical carcinogenesis. *Cancer* 47, 1055-1064.
36. Vineis, P. and Perera, F. (2000). DNA adducts as markers of exposure to carcinogens and risk of cancer. *Int J Cancer* 88, 325-328.
37. Tang, D., Phillips, D.H., Stampfer, M., Mooney, L.A., Hsu, Y., Cho, S., Tsai, W.Y., Ma, J., Cole, K.J., She, M.N. and Perera, F.P. (2001). Association between carcinogen-DNA adducts in white blood cells and lung cancer risk in the physicians health study. *Cancer Res* 61, 6708-6712.
38. Bonassi, S., Ugolini, D., Kirsch-Volders, M., Stromberg, U., Vermeulen, R. and Tucker, J.D. (2005). Human population studies with cytogenetic biomarkers: review of the literature and future perspectives. *Environ Mol Mutagen* 45, 258-270.
39. Albertini, R.J., Anderson, D., Douglas, G.R., Hagmar, L., Hemminki, K., Merlo, F., Natarajan, A.T., Norppa, H., Shuker, D.E., Tice, R., Waters, M.D. and Aitio, A. (2000). IPCS guidelines for the monitoring of genotoxic effects of carcinogens in humans. International Programme on Chemical Safety. *Mutat Res* 463, 111-172.
40. Mitelman, F., Mertens, F. and Johansson, B. (1997). A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nat Genet* 15 Spec No, 417-474.
41. Bonassi, S., Znaor, A., Norppa, H. and Hagmar, L. (2004). Chromosomal aberrations and risk of cancer in humans: an epidemiologic perspective. *Cytogenet Genome Res* 104, 376-382.
42. Bonassi, S., Lando, C., Ceppi, M., Landi, S., Rossi, A.M. and Barale, R. (2004). No association between increased levels of high-frequency sister chromatid exchange cells (HFCs) and the risk of cancer in healthy individuals. *Environ Mol Mutagen* 43, 134-136.
43. Bennett, D.A. and Waters, M.D. (2000). Applying biomarker research. *Environ Health Perspect* 108, 907-910.
44. Tate, A.D., Grummt, T., Tornqvist, M., Farmer, P.B., van Dam, F.J., van Mossel, H., Schoemaker, H.M., Osterman-Golkar, S., Uebel, C., Tang, Y.S. and et al. (1991). Biological and chemical monitoring of occupational exposure to ethylene oxide. *Mutat Res* 250, 483-497.
45. Schulte, P.A. and Waters, M. (1999). Using molecular epidemiology in assessing exposure for risk assessment. *Ann N Y Acad Sci* 895, 101-111.
46. Godschalk, R.W., Moonen, E.J., Schilderman, P.A., Broekmans, W.M., Kleijans, J.C. and Van Schooten, F.J. (2000). Exposure-route-dependent DNA adduct formation by polycyclic aromatic hydrocarbons. *Carcinogenesis* 21, 87-92.
47. Varkonyi, A., Kelsey, K., Semey, K., Bodell, W.J., Levay, G., Mark, E., Wain, J.C., Christiani, D.C. and Wiencke, J.K. (2005). Polyphenol associated-DNA adducts in lung and blood mononuclear cells from lung cancer patients. *Cancer Lett*
48. Wiencke, J.K., Kelsey, K.T., Varkonyi, A., Semey, K., Wain, J.C., Mark, E. and Christiani, D.C. (1995). Correlation of DNA adducts in blood mononuclear cells with tobacco carcinogen-induced damage in human lung. *Cancer Res* 55, 4910-4914.
49. Hamadeh, H.K., Amin, R.P., Paules, R.S. and Afshari, C.A. (2002). An overview of toxicogenomics. *Curr Issues Mol Biol* 4, 45-56.
50. Afshari, C.A., Nuwaysir, E.F. and Barrett, J.C. (1999). Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. *Cancer Res* 59, 4759-4760.
51. Nuwaysir, E.F., Bittner, M., Trent, J., Barrett, J.C. and Afshari, C.A. (1999). Microarrays and toxicology: the advent of toxicogenomics. *Mol Carcinog* 24, 153-159.
52. Cunningham, M.J., Liang, S., Fuhrman, S., Seilhamer, J.J. and Somogyi, R. (2000). Gene expression microarray data analysis for toxicology profiling. *Ann N Y Acad Sci* 919, 52-67.
53. Pennie, W.D. (2000). Use of cDNA microarrays to probe and understand the toxicological consequences of altered gene expression. *Toxicol Lett* 112-113, 473-477.
54. Fielden, M.R. and Zacharewski, T.R. (2001). Challenges and limitations of gene expression profiling in mechanistic and predictive toxicology. *Toxicol Sci* 60, 6-10.
55. Oberemm, A., Onyon, L. and Gundert-Remy, U. (2005). How can toxicogenomics inform risk assessment? *Toxicol Appl Pharmacol* 207, 592-598.

56. Aardema, M.J. and MacGregor, J.T. (2002). Toxicology and genetic toxicology in the new era of "toxicogenomics": impact of "-omics" technologies. *Mutat Res* 499, 13-25.
57. Simmons, P.T. and Portier, C.J. (2002). Toxicogenomics: the new frontier in risk analysis. *Carcinogenesis* 23, 903-905.
58. Orphanides, G. (2003). Toxicogenomics: challenges and opportunities. *Toxicol Lett* 140-141, 145-148.
59. Iannaccone, P.M. (2001). Toxicogenomics: "the call of the wild chip". *Environ Health Perspect* 109, A8-11.
60. Waters, M.D., Olden, K. and Tennant, R.W. (2003). Toxicogenomic approach for assessing toxicant-related disease. *Mutat. Res* 544, 415-424.
61. Hamadeh, H.K., Bushel, P.R., Jayadev, S., Martin, K., DiSorbo, O., Sieber, S., Bennett, L., Tennant, R., Stoll, R., Barrett, J.C., Blanchard, K., Paules, R.S. and Afshari, C.A. (2002). Gene expression analysis reveals chemical-specific profiles. *Toxicol Sci* 67, 219-231.
62. Waring, J.F. and Halbert, D.N. (2002). The promise of toxicogenomics. *Curr Opin Mol Ther* 4, 229-235.
63. Olden, K., Goehl, T.J. (2003). Inaugural Issue of EHP Toxicogenomics. *Environ Health Perspect* 111, A12.
64. Farr, S. and Dunn, R.T., 2nd (1999). Concise review: gene expression applied to toxicology. *Toxicol Sci* 50, 1-9.
65. Harries, H.M., Fletcher, S.T., Duggan, C.M. and Baker, V.A. (2001). The use of genomics technology to investigate gene expression changes in cultured human liver cells. *Toxicol In Vitro* 15, 399-405.
66. Thomas, P.S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc Natl Acad Sci U S A* 77, 5201-5205.
67. Liang, P. and Pardee, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257, 967-971.
68. Bustin, S.A. (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol* 29, 23-39.
69. Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995). Serial analysis of gene expression. *Science* 270, 484-487.
70. Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-470.
71. Pennie, W.D. and Kimber, I. (2002). Toxicogenomics; transcript profiling and potential application to chemical allergy. *Toxicol In Vitro* 16, 319-326.
72. Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E.L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14, 1675-1680.
73. Pennie, W.D., Woodyatt, N.J., Aldridge, T.C. and Orphanides, G. (2001). Application of genomics to the definition of the molecular basis for toxicity. *Toxicol Lett* 120, 353-358.
74. Ryder, M.I., Hyun, W., Loomer, P. and Haqq, C. (2004). Alteration of gene expression profiles of peripheral mononuclear blood cells by tobacco smoke: implications for periodontal diseases. *Oral Microbiol. Immunol* 19, 39-49.
75. Amundson, S.A., Do, K.T., Shahab, S., Bittner, M., Meltzer, P., Trent, J. and Fornace, A.J., Jr. (2000). Identification of potential mRNA biomarkers in peripheral blood lymphocytes for human exposure to ionizing radiation. *Radiat. Res* 154, 342-346.
76. Hellermann, G.R., Nagy, S.B., Kong, X., Lockey, R.F. and Mohapatra, S.S. (2002). Mechanism of cigarette smoke condensate-induced acute inflammatory response in human bronchial epithelial cells. *Respir. Res* 3, 22.
77. Fields, W.R., Leonard, R.M., Odom, P.S., Nordskog, B.K., Ogden, M.W. and Doolittle, D.J. (2005). Gene expression in normal human bronchial epithelial (NHBE) cells following in vitro exposure to cigarette smoke condensate. *Toxicol Sci* 86, 84-91.
78. Jorgensen, E.D., Dozmorov, I., Frank, M.B., Centola, M. and Albino, A.P. (2004). Global gene expression analysis of human bronchial epithelial cells treated with tobacco condensates. *Cell Cycle* 3, 1154-1168.
79. Sullivan, P.F., Fan, C. and Perou, C.M. (2006). Evaluating the comparability of gene expression in blood and brain. *Am J Med Genet B Neuropsychiatr Genet* 141, 261-268.
80. Spira, A., Schembri, F., Beane, J., Shah, V., Liu, G. and Brody, J.S. (2004). Impact of cigarette smoke on the normal airway transcriptome. *Chest* 125, 115S.
81. Spira, A., Beane, J., Shah, V., Liu, G., Schembri, F., Yang, X., Palma, J. and Brody, J.S. (2004). Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci U S A* 101, 10143-10148.
82. Forrest, M.S., Lan, Q., Hubbard, A.E., Zhang, L., Vermeulen, R., Zhao, X., Li, G., Wu, Y.Y., Shen, M., Yin, S., Chanock, S.J., Rothman, N. and Smith, M.T. (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.

83. Wang, Z., Neuburg, D., Li, C., Su, L., Kim, J.Y., Chen, J.C. and Christiani, D.C. (2005). Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. *Environ Health Perspect* 113, 233-241.
84. Lampe, J.W., Stepaniants, S.B., Mao, M., Radich, J.P., Dai, H., Linsley, P.S., Friend, S.H. and Potter, J.D. (2004). Signatures of environmental exposures using peripheral leukocyte gene-expression: tobacco smoke. *Cancer Epidemiol. Biomarkers Prev* 13, 445-453.
85. 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.
86. van Leeuwen, D.M., Gottschalk, R.W., van Herwijnen, M.H., Moonen, E.J., Kleinjans, J.C. and van Delft, J.H. (2005). Differential gene expression in human peripheral blood mononuclear cells induced by cigarette smoke and its constituents. *Toxicol Sci* 86(1), 200-210.
87. van Leeuwen, D.M., van Aken, B., Gottschalk, R.W.H., Vlietinck, R., Gielen, M., van Herwijnen, M.H.M., Maas, L., Kleinjans, J.C.S., and van Delft, J.H.M. (2006). Cigarette smoke-induced differential gene expression in blood cells from monozygotic twin pairs. *Carcinogenesis* (in press).
88. Van Leeuwen, D.M., Van Herwijnen, M., Pedersen, M., Knudsen, L., Kirsch-Volders, M., Sram, R., Staal, Y.C.M., Bajak, E., van Delft, J.H.M. and Kleinjans, J.C.S. (2005). Genome-wide differential gene expression in children exposed to air pollution in the Czech Republic. *Mutat Res* 600(1-2), 12-22.

Chapter 2

Differential gene expression in human peripheral blood mononuclear cells induced by cigarette smoke and its constituents

D.M. van Leeuwen
R.W.H. Gottschalk
M.H.M. van Herwijnen
E.J.C. Moonen
J.C.S. Kleinjans
J.H.M. van Delft

Department of Health Risk Analysis and Toxicology,
Maastricht University, Maastricht, the Netherlands

Toxicological Sciences 2005;86(1):200-10

Abstract

Currently, in molecular epidemiology studies, a wide range of methods is used to monitor early biological effects after exposure to xenobiotic agents. Gene expression profiling is considered a promising tool that may provide more sensitive, mechanism-based biomarkers. As a first step to obtain information on possible applicability of gene expression profiles as biomarker for early biological effects due to carcinogen exposure, we conducted *in vitro* studies on human peripheral blood mononuclear cells (PBMC) using cigarette smoke condensate (CSC) and a selection of its genotoxic constituents as model agents, using cDNA microarray technology to investigate modulated gene expressions. In independent experiments using cells from several donors, quiescent PBMC were exposed for 18 hours followed by gene expression analyses on a microarray containing 600 toxicologically relevant genes. The search for candidate biomarker genes was binomial: first for genes similarly responding to all agents, and second for agent-specific genes. Many genes were significantly deregulated by all compounds, but as the direction of deregulation frequently differed per agent, they are not useful as generic biomarker. CSC modulated the expression of many more genes than any of its constituents, with the largest effect in *SERPINB2*. The affected genes are involved in immune or stress response, but surprisingly no genes involved in DNA damage response were modulated, and only a few in DNA repair. In conclusion, several genes have been identified which may be potential biomarkers for population studies on early biological effects caused by cigarette smoke exposure, but no genes that represent a generic biomarker.

Introduction

Today, a range of biomonitoring methods is available for biomonitoring humans for the possible adverse effects caused by environmental exposure to hazardous compounds. Most of these methods have been developed to determine early biological effects due to exposure to xenobiotics. In general, these effects are detectable long before a diagnostic health effect appears [1,2], but generally different methods have to be applied for each health effect and for each agent or class of agents. It would be of profound benefit to molecular epidemiology to develop methods that would give (mechanism-based) information on several health effects simultaneously and that would be more suitable for monitoring effects at low exposure levels. Gene expression profiling is considered a promising tool to fulfill these needs. It is believed that the greatest potential for new biomarkers of early effect lies in toxicogenomics [3], which can be defined as a field of study that integrates toxicology and genomics and examines how the entire genome responds to toxicants or other environmental hazards [4,5].

Chemical carcinogenesis by environmental and occupational pollutants, as well as by life style factors like nutrition and tobacco smoke is of major interest in molecular epidemiology. Very likely, cigarette smoke is the main external source for human exposure to carcinogens [6]. Cigarette smoke contains over 4000 chemicals, including approximately 60 proven, probable and possible carcinogens according to IARC standards [7]. Cigarette smoke and a number of its chemical constituents are known to induce a broad range of (biomarkers of) genotoxic pre-carcinogenic effects, *e.g.* DNA adducts, micronuclei, chromosome aberrations and mutations, which have also been applied in molecular epidemiology studies using peripheral blood mononuclear cells (PBMC) as target cells [8,9].

In molecular epidemiology, target cells for toxic action are routinely not obtainable as highly invasive procedures would be required. However, for more than 10 years, evidence has been available that accessible cells, such as PBMC, can be used as surrogate target cells to monitor effects in target tissues [10,11]. Thus, these surrogate target cells seem also best suited for a biomonitoring method based on gene expression profiling. Up to now, only few studies on gene expression profiling in cells from exposed humans have been performed. Wu *et al.* [12] described altered gene expressions in PBMC as result of exposure to arsenic and demonstrated the potential of the technology to identify candidate biomarker genes. Recently, in a study on gene expression differences in peripheral leukocytes from human smokers versus non-smokers [13], genes were identified that significantly correlated with plasma cotinine levels, and their expression profiles accurately distinguish smokers from non-smokers. Importantly, another recent study on radiation exposure found *in vivo* gene expression patterns in blood to be similar to previously observed *in vitro* patterns in quiescent lymphocytes [35].

These human studies demonstrate the potential of gene expression profiling (in surrogate cells) in human biological monitoring studies. Furthermore, following an *in vitro* approach on gene expression profiling in human PBMC for the effects induced by ionizing radiation, Amundson *et al.* [15] have suggested that estimates of environmental radiation exposures may be provided by gene expression analyses in PBMC.

The aim of the present study was to investigate *in vitro* the gene responses induced by genotoxic carcinogens present in cigarette smoke using cDNA microarray technology, in a cellular model that best mimics the human *in vivo* situation as applied in biomonitoring studies (thus resting PBMC). Knowledge about these effects might aid in identifying potential biomarker genes for biomonitoring studies. Whether gene expression profiles are applicable for that, should be established in well designed future studies on exposed humans.

Materials and methods

Study design

The cell treatments were designed to represent the processes *in vivo* as much as possible. First, quiescent human PBMC were chosen rather than mitogen-activated cells. Human S9-mix was added to the treatment cultures. Since cigarette smoke compounds first pass the lung and only after absorption enter the blood and go through the liver, a mixture of S9 from human lung and liver was applied. Exceptions were made for treatments with H₂O₂ as S9 inactivates H₂O₂. Since smokers are chronically exposed to cigarette smoke or the compounds deposited in the lung, we choose to apply a long exposure period (18 h) rather than a pulse treatment (4-6 h). The 18 h period seems appropriate, as others have shown that, in general, deregulation of gene expression was larger and more genes were deregulated after approximately 24 hours compared to 4, 48 or 72 hours [14,15]. Furthermore, in many *in vitro* studies, a 18-24 h exposure period is common for investigating genotoxicity [16,17]. CSC-DNA adducts were measured to investigate the physiological relevance of the concentrations we applied *in vitro*, in comparison to adduct levels observed in smokers.

Test chemicals and materials

All chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA) unless stated otherwise. KCl and Mg₂Cl were from Merck AG (Darmstadt, Germany). Glucose-6-phosphate was from Boehringer Mannheim (Indianapolis, IN, USA), glucose-6-phosphate dehydrogenase and NADP were from Roche Diagnostics (Mannheim, Germany), L-glutamine, penicillin/streptomycin and fetal calf serum were from Gibco (Paisley, UK). Trizol and RPMI1640 medium were from Life Technologies (Breda, the Netherlands) and Lymphoprep from Nycomed (Oslo, Norway).

Chemical carcinogens used in this study were cigarette smoke condensate (CSC), benzo(a)pyrene (BaP), 4-(methylnitrosamino)-1-(3-pyridinyl)-1-butanone (NNK, Toronto Research, North York, Canada), 4-aminobiphenyl (4-ABP) and hydrogen peroxide (H₂O₂). BaP, NNK and 4-ABP are frequently investigated compounds representing three major classes of carcinogens present in cigarette smoke, *i.e.* polycyclic aromatic hydrocarbons, nitrosamines and aromatic amines. Hydrogen peroxide was included as a surrogate for oxidative stress induced by cigarette smoking. CSC was generated from 45 Kentucky 2R4F reference cigarettes on a 20-port Borgwaldt smoking machine under standardized ISO conditions, *i.e.* 35 ml/2 sec puff, 1 puff/min (kindly provided by Dr. K.

Rustemeier, INBIFO, Cologne, Germany) [18]. CSC stock solution was in DMSO at 100 mg total particulate matter (TPM) per ml.

Composition of cigarette smoke condensate

The composition of cigarette smoke condensate with respect to polycyclic aromatic hydrocarbons was analyzed by high performance liquid chromatography (HPLC) according to the NIOSH protocol issue 3 [19,20]. The average concentration of BaP was 21.27 ng/ml (83 μ M). NNK and 4-ABP concentrations were determined by standard GC/MS (mass resolution 10000) and were 161.0 ng/ml (770 μ M) $\pm$ 11% and 57.33 ng/ml (338 μ M) $\pm$ 32.2% respectively.

Collection of blood samples

Buffy coats of 500 ml whole blood from 7 healthy non-smoking donors (age 20-50 years) were randomly obtained from Sanquin Blood Bank Limburg in Maastricht. Donors gave their written informed consent. One buffy coat was sufficient for assessing gene expression profiles of 2 agents at three concentrations and a concomitant vehicle control.

PBMC isolation

Human PBMC were isolated from buffy coat through gradient centrifugation using Lymphoprep in Leucosep filter tubes (Greiner Bio-One, Frickenhausen, Germany) according to the manufacturer's protocol. Finally, cells were resuspended in RPMI 1640 medium, supplemented with 10% heat-inactivated fetal calf serum, L-glutamine (2 mM) and penicillin/streptomycin (penicillin 50 U/ml and streptomycin 50 μ g/ml) at $6-8 \times 10^6$ cells/ml. Cells were quiescent, meaning that no mitotic stimulation was applied.

Metabolic activation system

S9 fractions of human liver and human lung were used for metabolic activation of the carcinogens. Pooled human liver S9 fraction from 15 donors (males and females) was purchased from In Vitro Technologies (Baltimore, MA, USA). Human lung S9 fractions were prepared from healthy tissue of 4 different patients who had undergone lung tumor surgery at the Maastricht Academic Hospital. Liver and lung S9 fractions were mixed and each covering 15% of the total S9 mix, which furthermore consisted of 32 μ l 1M KCl, 31 μ l 0.25 M $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$, 24,3 μ l 0.2 M glucose-6-phosphate, 26 μ l 140 U/ml glucose-6-phosphate dehydrogenase and 97,4 μ l 0.04 M NADP with 204,5 μ l MQ and 292 μ l 1x PBS per ml [21].

Toxicity testing

Toxicity levels of the agents were determined by flow cytometry analysis. Briefly, PBMC were exposed for 18 h to the test agents in wide concentration ranges (CSC from 18 to 500 μ g/ml, BaP from 0.1 to 100 μ M, NNK from 2 to 2000 μ M, 4-ABP from 0.1 to 100 μ M, and H_2O_2 from 3×10^{-5} to 0.3 % (v/v)), all in the presence (10% v/v) and absence of S9-

mix, and at least in duplicate. After exposure propidium iodide was added (20 $\mu\text{g/ml}$). Propidium iodide intercalates with DNA and only stains dead cells because of the membrane impermeability of viable cells. Samples were analyzed on a FACSort flow cytometer (Becton Dickinson, Heidelberg, Germany). Based on forward-sideward scatter plot, 10,000 PBMC were analyzed for cytotoxicity (number of stained cells as percentage of total cell count).

In vitro exposure

PBMC were exposed for 18 hours continuously to three concentrations of each agent at 37°C in a total volume of 5 ml ($6\text{--}8 \times 10^6$ cells/ml) in 15 ml polypropylene conical tubes (Greiner BioOne). CSC, BaP, 4-ABP and NNK exposures took place in the presence of S9 mix (10% v/v), H_2O_2 exposure without metabolic activation (S9 inactivates H_2O_2). Control cells were exposed to the vehicle solvents only, *i.e.* DMSO (CSC, BaP, 4-ABP and NNK) or PBS (H_2O_2). Control cells always originated from the same buffy coat as the treated cells. After 18 hours, tubes were chilled on ice and centrifuged at 250x g for 15 minutes. Cells were lysed and homogenized in 1 ml Trizol[®] / 10^7 cells for RNA stabilization and subsequent RNA isolation. Independent treatments with cells from different donors were performed twice for the low and medium concentrations of the individual agents, and three times for the high concentrations of the individual agents and for all concentrations of CSC. We believe that by investigating effects in cells from three different donors, a reasonable estimation is obtained of the gross effects that may occur in the human population.

Total RNA isolation

Total RNA was isolated by a combination of two methods. First, total RNA was isolated using the Trizol[®] method, according to the manufacturer's protocol. Next, RNA was purified using RNeasy[®] Mini Kit (Qiagen) according to the manufacturer's protocol. RNA was dissolved in DEPC-treated, RNase-free, water. Purity was tested spectrophotometrically and considered suitable for further processing at 260/280 ratios of >1.7 . Integrity was tested by lab-on-a-chip technology on the BioAnalyzer (Agilent, Palo Alto, CA, USA) and RNA was considered to be intact when showing two distinct bands for the 18S and 28S ribosomal RNA.

cDNA labeling and microarray hybridization

Reverse transcription was carried out on 10 μg of total RNA with minor modifications according to the TIGR protocol for aminoallyl labeling for microarrays [22]. Finally, the test and accompanying control samples were mixed and dried *in vacuo*. As technical replicates, each test sample was hybridized twice against its concomitant vehicle control; once Cy5 labeled (control Cy3) and once Cy3 labeled (control Cy5). Samples were hybridized under a cover slip to spotted glass PHASE-1 Human 600 cDNA microarrays (PHASE-1 Molecular Toxicology, Santa Fe, NM, USA) according to the manufacturer's protocol. These microarrays contain 597 toxicologically relevant human genes in quadruplicate (in a 2x2 design) representing several gene categories, such as apoptosis, cell

proliferation, cell cycle, DNA damage and repair, inflammation, metabolism, oxidative stress, peroxisome proliferators, transport and cell environment. Each gene is represented on the array in quadruplicate, which adds to a lower inter and intra assay variability. Samples were dissolved in 30 μ l hybridization buffer containing 50% formamide, 5x SSC; 0.1% SDS and 100 μ g/ml salmon sperm DNA. Samples were denatured for 5 min at 95°C, centrifuged at maximal speed for 3 minutes and kept at 70°C until hybridization. Then, 25 μ l was placed on the microarrays and incubated in hybridization chambers (Corning Inc., Corning, NY, USA) overnight (18 hours) in a water bath at 42°C. After hybridization, slides were dipped in 2x SSC at 34°C to remove the cover slips. Thereupon, slides were washed in 2x SSC/0.1% SDS, 0.1x SSC/0.1% SDS and 0.1x SSC for 5 minutes each. Finally, slides were washed in 0.1x SSC at 32°C and spun to dry. Microarrays were scanned within 3 days and stored in a dark, dry environment.

Image and data analysis

Microarrays were scanned on the Affymetrix GMS 418 scanner (Santa Clara, CA, USA) at a fixed photomultiplier gain of 65 and a variable laser power, such that no spots were saturated. Cy5 and Cy3 fluorescent signals were scanned in separate channels. TIFF were imported pairwise (Cy5 and Cy3) in ImaGene 5.0 (BioDiscovery, Marina del Rey, CA, USA) for image analysis, resulting in raw data files with average pixel intensities for every spot and its local background. Poor spots were manually flagged (these are spots that are not well processed by ImaGene, generally due to severe blurring or contamination with a dust particle). Raw data files were imported in GeneSight 4.0 (BioDiscovery) for further data analysis. First, for every spot local background was subtracted, then manually flagged poor spots and spots with low expression levels (<5 mean pixel intensities) were eliminated. After log₂ transformation of the remaining data, LOWESS normalization was performed on all genes with a smoothing factor of 0.2, and the difference of each test relative to their concomitant vehicle control samples was calculated. Finally, replicate spots and arrays were averaged for each gene, omitting outliers with >2 standard deviations.

In order to identify differentially expressed genes, t-tests were performed on the difference data of all genes for each concentration of each agent relative to a set of 4 self-self hybridizations. In a self-self hybridization two aliquots of the same RNA sample were hybridized against each other, which results in the estimation of the experimental variation. T-tests were performed with Holm's p-value adjustment, to reduce the number of false positives, with significance levels of $p < 0.001$, $p < 0.01$ and $p < 0.05$. T-testing was also done without Holm's adjustment. Also, for every agent an ANOVA-test was performed including all concentrations and the self-self hybridizations, to reveal genes with dose-response relations. For these analyses, the same restrictions regarding the Holm's adjustments and p-values were maintained as described above. As this is an initial exploratory study, we did not target the aspect of inter-individual variations; this will be the subject of future *in vivo* investigations. Therefore, the data of all donors were pooled for each concentration and each agent.

Annotation and ontology analyses were done using the Database for Annotation, Visualization and Integrated Discovery (DAVID, <http://apps1.niaid.nih.gov/David>) [23]. Analysis of overrepresentation of specific biological pathways by selected genes

compared to all genes present on the microarrays was conducted using the online tool Expression Analysis Systematic Explorer (<http://david.niaid.nih.gov/david>) [24].

Quantitative real-time PCR

Quantitative real-time PCR (qPCR) was performed to quantitate mRNA levels of a selection of genes in order to verify expression changes resulting from the microarray experiments. cDNA was prepared from 1 µg total RNA using random hexamer primers and SuperScript II reverse transcriptase (Invitrogen, Life Technologies) and used for real-time qPCR with the qPCR Core Kit for SYBR Green I, according to the manufacturers instructions (Eurogentec, Seraing, Belgium). All PCR reactions were performed in duplicate. β -actin was used as reference in order to normalize expression levels and to quantitate changes in gene expressions between the control and treated samples. The qPCR was run on the ABI Prism 7500 Sequence Detector (Applied BioSystems, Foster City, USA); 2 min at 50°C, 10 min at 95°C and 40 cycles of 95°C for 15 seconds and 60°C for 1 min.

The following forward and reverse primers (OPERON, 5'-3' sequences):

IL1 β

CTGAGCTCGCCAGTGAAATG (forward); TTTAGGGCCATCAGCTTCAA (reverse),

IL6

TCCAGGAGCCCAGCTATGAA (forward); GAGCAGCCCCAGGGAGAA (reverse),

TGM2

GGGCTCGGCCAAGTTCAT (forward); TCTAGAAGGATCAGGCAGATGTCTAG (reverse),

SERPINE2

GGGCTTTATCCTTTCCGTGTA (forward); TTAGCTTTTCACGCAAGTACATCAT (reverse)

β -actin

CCTGGCACCCAGCACAAAT (forward) and GCCGATCCACACGGAGTACT (reverse).

Dissociation curve analysis was performed and 'no template controls' were analyzed to check for non-specific products in the reaction. For each sample, the quantity was derived from $\Delta Ct = Ct_{(\text{target gene})} - Ct_{(\beta\text{-actin reference})}$. These values were log2 transformed and the difference of each test relative to their concomitant vehicle control samples was calculated.

DNA-adduct analysis

Aromatic DNA adduct levels were assessed by ^{32}P -postlabeling. DNA was isolated from the CSC Trizol samples of three donors according to the manufacturer's instructions. The ^{32}P -postlabeling procedure was conducted as originally described by Reddy and Randerath [25] with minor modifications as described by Godschalk *et al.* [26]. Two DNA samples with known levels of modification (1 adduct per 10^7 and 10^8 nucleotides) were included in the analyses for quantitation purposes. Radiolabeled adduct nucleotide bi-phosphates were separated by two dimensional chromatography on PEI-cellulose sheets (Machery Nagel, Germany), using solvent systems described by Godschalk *et al.* [26]. Chromatograms were visualized and quantified using a FLA-3000 Fuji Phosphor Imager (Fuji, Paris, France) with AIDA/2D Densitometry software.

Results

Toxicity tests

Concentration levels for the gene expression studies were based on cytotoxicity assessment by flow cytometry analysis of treated cells. Except for H₂O₂ in the absence of S9, none of the agents showed substantial toxicity at the tested concentrations as cytotoxicity levels were always <10%. For H₂O₂, 1*10⁻³ % (v/v) showed a cytotoxicity of 34% (median of 4 donors, ranging from 17% to 83%) whereas 3*10⁻⁴% (v/v) showed 23% toxicity (4 donors, ranging from 4% to 68%) and 1*10⁻⁴% showed 3% toxicity (average 2 donors). The final concentrations chosen for application in the gene expression analyses were: CSC 30, 100 and 300 µg/ml, BaP 3, 10 and 30 µM, 4-ABP 10, 30 and 100 µM, NNK 100, 300 and 100 µM and H₂O₂ 0.0001, 0.0003 and 0.001%.

Gene expression analyses

Human PBMC were *in vitro* exposed to CSC, BaP, 4-ABP, NNK or H₂O₂ at three concentrations followed by analyses of altered mRNA levels by cDNA microarrays. Differentially expressed genes per agent and concentration were identified by t-testing against self-self hybridizations and for concentration-related effects of each agent by ANOVA of the three concentrations combined with self-self hybridizations. The numbers of significant gene expression modulations in the various tests are shown in Table 1.

Table 1. Numbers of differentially expressed genes per concentration for each agent at different p-values by t-tests or ANOVA.

Concentration	p-value [#]	CSC	BaP	NNK	4-ABP	H ₂ O ₂
Low (t-test)	0.01 +H	7 (34)	3 (16)	3 (39)	5 (30)	0 (19)
	0.05 +H	8 (68)	4 (53)	6 (93)	5 (68)	1 (46)
Medium (t-test)	0.01 +H	7 (53)	1 (38)	1 (11)	5 (22)	0 (18)
	0.05 +H	12 (99)	5 (73)	1 (38)	6 (53)	1 (51)
High (t-test)	0.01 +H	30 (136)	0 (22)	0 (23)	1 (16)	0 (19)
	0.05 +H	47 (196)	2 (62)	2 (60)	1 (54)	1 (74)
All* (ANOVA)	0.01 +H	28 (103)	2 (23)	1 (29)	0 (14)	0 (10)
	0.05 +H	38 (137)	4 (52)	3 (59)	3 (32)	0 (36)
Unions**	0.01 +H	39 (169)	5 (68)	5 (78)	8 (51)	0 (49)
	0.05 +H	56 (260)	12 (169)	9 (178)	9 (143)	3 (138)

* All concentrations combined with self-self hybridizations for establishment of dose-response relations.

** All genes differentially expressed by an agent in any of the above tests.

+H = test with Holm's adjustment of the p-value. Values in brackets represent number of genes differentially expressed without Holm's adjustment.

In order to identify candidate biomarker genes suitable for human biominotoring studies, two approaches were followed. We first looked for genes that were modulated by all agents, as these genes when affected in the same direction by all agents, may be specific for genotoxic stress to PBMC without discriminating the type of agent. Figure 1 shows an adaptation to the three-ring Venn diagram, presenting the numbers of genes that were differentially expressed by CSC and one or more of the other agents. The intersection of all areas, the centre of the diagram, contains 16 genes that were differentially expressed by all agents (at $p < 0.05$). Interleukin-6 (*IL6*), platelet/endothelial cell adhesion molecule-1 (*PECAM1*) and spermidine/spermine N1-acteyltransferase (*SAT*) were most significantly modulated ($p < 0.01$ without Holm's adjustment). A heat map presenting expression differences for all 16 genes is shown in Figure 2. None of the genes appears to be affected in the same direction by all agents, but the direction of deregulation differs per agent. Therefore, these genes appear not useful as general biomarker for exposure to genotoxic carcinogens.

Figure 1. Diagram of genes modulated by both CSC and one or more of the individual agents. Numbers in specific areas represent numbers of differentially expressed genes by CSC and the specific agent (unions of all tests at $p < 0.05$ without Holm's adjustment; see Table 1). Gene names in the middle section indicate the 16 genes that were differentially expressed by all agents at $p < 0.05$, and bold and underlined names indicate genes differentially expressed at $p < 0.01$.

Figure 2. Heatmap presenting expression changes for all 16 genes that were differentially expressed by all agents (at $p < 0.05$; see Figure 1).

Second, we searched for the most sensitive genes per agent, meaning that they responded at low or medium concentration and at high significance ($p < 0.05$ with Holm's p-value adjustment; see Table 1). This should yield agent-specific genes. The differential expression for these genes is shown in Figure 3. As is clear, CSC caused the largest effect on the gene expression profile. The effects induced by the other agents are less pronounced. For BaP, NNK and H_2O_2 often a less than 2-fold induction or suppression is observed. Only 4-ABP caused clear concentration-related effects for several genes. Furthermore, the data show that the effects caused by CSC are far more complex and can not be accounted for by a simple addition of the results of the other individual agents.

Figure 3. Differential expression of all genes that were significantly modulated at low and/or medium concentrations ($p < 0.05$ with Holm's adjustment; see Table 1). Error bars represent standard errors. An asterisk indicates if a gene was significantly differentially expressed at the low concentration.

For the agent-specific genes that showed a more than 2-fold difference ($p < 0.05$ with Holm's adjustment, Figure 3), information about molecular function and biological process is shown in Table 2.

In order to validate expression changes resulting from the microarray experiments, real-time qPCR was performed to quantify mRNA levels for the main agent-specific deregulated genes, i.e. *IL1 β* , *IL6*, *TGM2* and *SERPINB2* for CSC, *SERPINB2* for BaP, and *IL1 β* , *IL6* for 4-ABP (Figure 4). All these RT-PCR data confirmed the microarray data.

Figure 4. Gene expression analyses by real-time qPCR to validate of the most relevant differentially expressed genes identified by the microarray study for CSC, BaP and 4-ABP (see Figure 3). Error bars represent standard errors of the mean.

Table 2. Overview of the most relevant differentially expressed genes. Detailed information is only provided for the agent-specific genes with an expression difference of ≥ 2 -fold (>1 or <-1 at a $\log_2$ -scale of Figure 3).

GenBank	Abbreviation	Gene name	Effect on expression	Molecular function and biological process ¹⁾	Main function ²⁾
Z11559	ACO1	aconitase 1, soluble	CSC: $\uparrow$ 4-ABP: $\downarrow$	RNA binding; aconitate hydratase activity; iron ion binding; lyase activity; negative regulation of translation	cellular iron homeostasis
NM_004849	APG5L	APG5 autophagy 5-like (S. cerevisiae)	$\downarrow + \uparrow$	apoptosis; autophagy	expressed at high levels in apoptosis
L19871	ATF3	activating transcription factor 3	$\uparrow$	regulation of transcription, DNA-dependent; transcription corepressor activity	transcription regulation
X03663	CSF1R	colony stimulating factor 1 receptor, formerly McDonough feline sarcoma viral (v-fms) oncogene homolog	$\downarrow$	ATP binding; antimicrobial humoral response (sensu Vertebrata); cell proliferation; development; macrophage colony stimulating factor receptor activity; protein amino acid phosphorylation; transmembrane receptor protein tyrosine kinase signaling pathway	receptor for CSF1 that controls production, differentiation, and function of macrophages
AF043340	CXCL2	chemokine (C-X-C motif) ligand 2	$\uparrow + \downarrow$	G-protein coupled receptor protein signaling pathway; chemokine activity; chemotaxis; immune response	regulation leukocytes and immune system
14328	ENO1	enolase 1, (alpha)	$\downarrow$	DNA binding; glycolysis; lyase activity; magnesium ion binding; negative regulation of cell growth; negative regulation of transcription from Pol II promoter; phosphopyruvate hydratase activity; protein binding; transcriptional repressor activity	glycolytic enzyme
L20046	ERCC5	excision repair cross-complementing rodent repair deficiency, complementation group 5	$\downarrow$	endodeoxyribonuclease activity; single-stranded DNA binding; transcription-coupled nucleotide-excision repair	excision repair of UV-induced DNA damage
AF309489	HAMP	hepcidin antimicrobial peptide	$\uparrow + \downarrow$	defense response to fungi; hormone activity; immune response; xenobiotic metabolism	component of innate immunity
X02532	IL1B	interleukin 1, beta	CSC: $\uparrow + \downarrow$ 4-ABP: $\downarrow$	antimicrobial humoral response (sensu Vertebrata); apoptosis; cell proliferation; cell-cell signaling; immune response; negative regulation of cell proliferation; regulation of cell cycle; signal transduction	mediator of inflammatory response

XM_004777	IL6	interleukin 6 (interferon, beta 2)	CSC: ↑ + ↓ 4-ABP: ↓	acute-phase response; cell proliferation; cell surface receptor linked signal transduction; cell-cell signaling; cytokine activity; development; humoral immune response; regulation of cell proliferation	immune regulator
Y00787	IL8	interleukin 8	↓	G-protein coupled receptor protein signaling pathway; angiogenesis; calcium-mediated signaling; cell cycle arrest; cell motility; cell-cell signaling; chemokine activity; immune response; induction of positive chemotaxis; interleukin-8 receptor binding; intracellular signaling cascade; negative regulation of cell proliferation; neutrophil activation and chemotaxis; regulation of cell adhesion; regulation of retroviral genome replication	mediator inflammatory response
X92720	PCK2	phosphoenolpyruvate carboxykinase 2 (mitochondrial)	↓ + ↑	GTP binding; gluconeogenesis; lyase activity; manganese ion binding; phosphoenolpyruvate carboxykinase (GTP) activity	enzyme in gluconeogenesis
XM_001734	PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	↑ + ↓	cell motility; cyclooxygenase pathway; oxidoreductase activity, peroxidase activity; prostaglandin-endoperoxide synthase activity; regulation of inflammatory response	prostaglandin biosynthesis
M55421	SCP2	sterol carrier protein 2	↑	lipid binding and transport; steroid biosynthesis; sterol carrier activity	thought to be involved in intracellular lipid transfer
M18082	SERPINE2	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 2	↑	anti-apoptosis; plasminogen activator activity; serine-type endopeptidase inhibitor activity	primary regulator of plasminogen activation
J04765	SPP1	secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)	↓	T-helper 1 type immune response; anti-apoptosis; cell-cell signaling; cell-matrix adhesion; cytokine activity; growth factor activity; immune cell chemotaxis; integrin binding; negative regulation of bone mineralization; ossification; positive regulation of T-cell proliferation; protein binding; regulation of myeloid blood cell differentiation	anchoring of osteoclasts to the mineral of bone matrix
M98478	TGM2	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	↓	acyltransferase activity; calcium ion binding; peptide cross-linking	catalyzes protein crosslinking

¹⁾ Gene Ontology Terms from Gene Ontology at www.ncbi.nlm.nih.gov/LocusLink

²⁾ Derived from OMIM or RefSeq (via www.ncbi.nlm.nih.gov/LocusLink)

DNA adduct analysis

In addition to the gene expression analyses, PAH-DNA adduct levels were measured in the CSC treated samples and the controls by ^{32}P -postlabeling. In the exposed samples, elevated numbers of adducts were measured when compared to the control samples, showing a concentration-effect relationship (Table 3).

Table 3. PAH-DNA adduct levels in the CSC treated PBMC samples and controls as measured by ^{32}P -postlabeling (mean $\pm$ standard error of the mean)

CSC concentration	Adducts/ 10^8 nucleotides
Control (DMSO)	0.64 ± 0.25
Low (30 $\mu\text{g/ml}$)	1.39 ± 0.18
Medium (100 $\mu\text{g/ml}$)	1.55 ± 0.28
High (300 $\mu\text{g/ml}$)	3.56 ± 0.82

Discussion

Our scope was to examine *in vitro* in a cellular model that best mimics the human *in vivo* situation as applied in biomonitoring studies (thus resting PBMC), the gene responses induced by genotoxic carcinogens present in cigarette smoke, by using cDNA microarray technology. Knowledge about these effects may aid in identifying potential biomarker genes for which expression changes can be applied as a biomarker of early effect in biomonitoring studies. For this purpose, we evaluated effects of cigarette smoke condensate and a selection of its major carcinogenic constituents on human quiescent PBMC *in vitro*. Here we report the induction of differential gene expression in this model and our approach to identify candidate biomarker genes. The present study demonstrates that gene expression in PBMC can be significantly modulated *in vitro* by exposure to genotoxic chemicals. By exposing human PBMC *in vitro* to CSC, BaP, NNK, 4-ABP and H_2O_2 , we identified genes that were differentially expressed by all agents (the intersection for all agents from Figure 1), as well as differentially expressed genes for each agent individually (differentially expressed at the two lowest concentrations; see Figure 3). We hypothesize that these genes are most promising for future biomonitoring purposes in human population studies.

Genes that were modulated by all agents can be considered to represent a general response to genotoxic stress, as the genes responded to each agent investigated in this study. Thus, hypothetically, monitoring the expression levels for these (group of) genes in a biomonitoring study may discriminate persons exposed to toxic agents from those that are not exposed, without discriminating the type of toxicant. This requires though, that the deregulation must be robust and in the same direction for all agents. However, the heat map presenting expression changes for all 16 genes shows that this is not the case (Figure 2). None of the genes appeared to be consistently affected in the same direction by all agents, but the direction of deregulation rather ap-

peared to be agent-specific. Therefore, we conclude that these genes are not useful as generic biomarker for exposure to genotoxic carcinogens in biomonitoring studies.

The genes differentially expressed at the lowest and/or medium concentrations for any agent can be regarded as the most relevant agent-specific genes. Of all agents, CSC induced alterations in the largest number of genes (Table 1; Figure 3). For a number of genes, dose-related effects from low up to high concentrations were found (e.g. for *TGM2*, *ATF3*, *ENO1* and *VEGF*), whereas for other genes constant expression changes were observed (*CSF1R*, *SSP1* and *SERPINB2*). The latter may be due to saturation and would thus require exposure to lower concentrations of CSC to avoid this effect. For several genes, modulation of gene expression did not show a clear concentration-response relationship. Especially the highest concentration deviated, which in some cases caused a reversal of the expression change (e.g. for *IL1 β* , *IL6* and *PTGS2*). As this has been independently observed for PBMC from several donors, this does not reflect inaccuracy of the method. Furthermore, the qPCR analyses confirm the microarray data (compare Figure 3 and 4). Presumably, the change in the direction of deregulation is due to pre-cytotoxic processes that might have been induced by the highest CSC concentration, that were not yet apparent in our cytotoxicity tests.

DNA adduct levels in CSC exposed cells showed a clear concentration-related effect. The DNA adduct data confirm that the lowest CSC concentration was physiologically relevant, since the adduct level at this concentration is comparable to that in lymphocytes from smokers [27]. Therefore, induced gene expression profiles can be considered physiologically relevant as well.

A response of the PBMC to CSC or BaP exposure that could have been expected in advance, concerns the differential expression of genes regulated by the aryl hydrocarbon receptor (AhR). AhR is known for its induction of several genes, such as cytochrome P450 1A1, 1A2 and 1B1, in response to exposure to these agents [28]. In mitogen-activated lymphocytes, indeed *CYP1A1* and *CYP1B1* are inducible by TCDD, a very potent inducer of AhR activity [29]. However, in quiescent lymphocytes this induction does not seem to occur, since long term presence of dioxin in the human body does not result in an increase in AhR pathway responsiveness [29]. In our study, differential gene expression of AhR related genes could not be demonstrated, which indicates that this AhR signalling pathway is not very active in quiescent PBMC, and which agrees with the population study by Landi *et al.* [29]. In a comparable gene expression study in HepG2 cells, the AhR related genes were activated by BaP [30], thus showing that the technology is sufficiently sensitive. Furthermore, in our study DNA damage response and DNA repair related genes hardly showed modulations of their expression. None of the DNA damage response genes regulated by p53 and induced in exposed HepG2 cells, such as *BAX*, *CDKNA2* and *GADD45* [30], were significantly affected in PBMC. CSC, BaP and NNK exposed cells did, however, show one significantly deregulated DNA repair gene, namely *XRCC1*, *XRCC5* and *ERCC5*, respectively (Figure 3). In all cases, however, a down-regulation was observed where up-regulation would be expected, and the effects were marginal without obvious concentration-response relationships.

From the subset of genes that most sensitively responded to the exposures, we specifically focused on the genes that showed a ≥ 2 fold ($-1 \leq$ expression difference ≤ 1 ; see Figure 3) up or down regulation. These are *CSF1R*, *CXCL2*, *SPP1*, *SERPINB2*, *TGM2*, *PTGS2*, *ATF3*, *ACO1*, *ENO1* (for CSC), *SERPINB2*, *PCK2* (for BaP), *ERCC5*, *APG5L* (for NNK), *IL1 β* , *IL6*, *IL8*, *HAMP*, *SCP2* and *ACO1* (for 4-ABP). Annotation

analyses by DAVID (<http://david.niaid.nih.gov/david>) showed that 8 of these genes were involved in or associated with immune response, 10 in metabolism, 4 in apoptosis, 5 in cell proliferation and 5 in signal transduction (see also Table 2). Analysis of overrepresentation of specific biological pathways in this selection of genes compared to all genes present on the microarrays by EASE (<http://david.niaid.nih.gov/david>), showed that biological processes involved in immune and defense response were most significantly affected (Fisher Exact p-values 0.00031 and 0.00036). This emphasizes that *in vitro* exposure of PBMC to these agents particularly activates biological pathways that are involved in defense response implicating genes of the immune system. Whether this relates to the fact that we used PBMC which are involved in immune response and therefore may be more responsive in these pathways, or alternatively that many genotoxicants are also immunotoxicants [31,32], remains yet unanswered. The absence of modulation of DNA damage response pathways may indicate that rather non-specific responses were induced, but might also imply that DNA damage response pathways are not induced because quiescent cells were exposed. In order to further investigate this matter, it would be prudent to follow up our experiments with a selection of other classes of toxic agents (e.g. non-genotoxic agents) and with mitogen-stimulated PBMC.

EASE analyses per agent, showed that CSC significantly affected the immune/defense response (Fisher Exact p-value), and that 4-ABP affected immune/defense response and regulation of cell proliferation, whereas for BaP, NNK and H₂O₂ no specific biological processes were significantly affected. It is remarkable that 4-ABP was consistently shown to influence expression of immune response related genes, and that the change was in the opposite direction compared to CSC. It is likely that the suppressed expression has repercussions for the functionality of the cells. Yet, immunotoxicity of 4-ABP has not been reported. As such, these observations warrant a study of potential immunotoxicity of this agent. It is noteworthy that BaP failed to influence expression of immune response related genes, whereas BaP is a known immunotoxic agent [33]. BaP possibly induces immunotoxicity through the Ah receptor, which is hardly expressed in lymphocytes, as discussed above for the absence of the AhR response.

The genes that were differentially expressed by a specific agent, especially at the lowest concentrations, can be considered to represent agent-specific responding genes (Figure 3). Thus, hypothetically, monitoring the expression levels for these (groups of) genes in a biomonitoring study may discriminate persons exposed to the specific agents from those that are not exposed. Taking into account that only CSC and 4-ABP caused concentration-related effects for several genes, possibly only gene expression analyses related to exposure to these agents might be sufficiently achievable.

Taking into account the microarray and qPCR data, we believe that *IL1 β* , *IL6*, *TGM2* and *SERPINB2* are the most promising potential biomarker genes for effects caused by cigarette smoke exposure, where the effects on *SERPINB2* may be ascribed to BaP or possibly other polycyclic aromatic hydrocarbons present in cigarette smoke. As mentioned earlier, it seems not possible to identify (profiles of) genes that can be used as general biomarker for genotoxic agent exposure, but before reaching a final conclusion on this, these *in vitro* experiments should be followed up by studies on humans that are heavily exposed to cigarette smoke. Therefore, our current study must be seen as an initial, exploratory step in the development of gene expression profiles as biomarker of early effect.

To date, several other *in vitro* and *in vivo* studies on gene expression profiling in PBMC following exposure to a carcinogenic agents like cigarette smoke and ionizing radiation have been published [12,13,15,34,35]. These studies show that expressions of genes involved in inflammation or immune response were frequently modulated. Also, a study has been published on acute inflammatory response in human bronchial epithelial cells, a target cell for cigarette smoke exposure, which describes increased expression of *IL1 β* , *IL6* and *IL8* induced by *in vitro* exposure to CSC [36]. This substantiates our observations in PBMC *in vitro*. For *SERPINE2*, which has plasminogen activation regulation as its main function, induction by TCDD exposure has been observed in monocytic U937 cells [37]. *TGM2* catalyzes protein crosslinking and its activity has been shown to be increased in apoptotic hepatocytes [38] suggesting a role for the gene in regulation of cell death. However, we can not address the relevance of this for our findings.

In our study, we used PBMC as surrogate targets cells for effects that may occur in the true targets organs *in vivo*, as this is also the approach in biomonitoring studies of human populations. In human volunteers it is *a priori* not possible to investigate biological effects in cells from target organs like lung and liver [10,11]. A critical question is whether effects in PBMC at gene expression level are comparable to that in target organs. To our knowledge, this has not yet been investigated in humans or animals.

In summary, our study demonstrates that cigarette smoke condensate and its constituents caused gene expression changes in human PBMC *in vitro*. The identified differentially expressed genes have the potential to become marker genes for population studies on biological effects caused by cigarette smoke exposure. However, these findings need to be substantiated *in vivo*.

Acknowledgements

The authors thank Dr. K. Rustemeier (INBIFO, Cologne, Germany) for kindly generating and providing the cigarette smoke condensate, Dr. B. Schutte (Department of Genetics, Maastricht University, Netherlands) for assisting on flow cytometry and Prof. dr. H. van Loveren (Department of Health Risk Analysis and Toxicology, Maastricht University, Netherlands) for his comments. This study was supported by the Flanders Environment and Health Study.

References

1. Bonassi, S. and Au, W.W. (2002). Biomarkers in molecular epidemiology studies for health risk prediction. *Mutat. Res* 511, 73-86.
2. van Delft, J.H., Baan, R.A. and Roza, L. (1998). Biological effect markers for exposure to carcinogenic compound and their relevance for risk assessment. *Crit. Rev. Toxicol* 28, 477-510.
3. Toraason, M., Albertini, R., Bayard, S., Bigbee, W., Blair, A., Boffetta, P., Bonassi, S., Chanock, S., Christiani, D., Eastmond, D., Hanash, S., Henry, C., Kadlubar, F., Mirer, F., Nebert, D., Rapport, S., Rest, K., Rothman, N., Ruder, A., Savage, R., Schulte, P., Siemiatycki, J., Shields, P., Smith, M., Tolbert, P., Vermeulen, R., Vineis, P., Wacholder, S., Ward, E., Waters, M. and Weston, A. (2004). Applying new biotechnologies to the study of occupational cancer--a workshop summary. *Environ. Health Perspect* 112, 413-416.
4. Cunningham, M.L., Bogdanffy, M.S., Zacharewski, T.R. and Hines, R.N. (2003). Workshop overview: use of genomic data in risk assessment. *Toxicol. Sci* 73, 209-215.
5. Waters, M.D., Olden, K. and Tennant, R.W. (2003). Toxicogenomic approach for assessing toxicant-related disease. *Mutat. Res* 544, 415-424.
6. Villard, P.H., Seree, E.M., Re, J.L., De Meo, M., Barra, Y., Attolini, L., Dumenil, G., Catalin, J., Durand, A. and Lacarelle, B. (1998). Effects of tobacco smoke on the gene expression of the Cyp1a, Cyp2b, Cyp2e, and Cyp3a subfamilies in mouse liver and lung: relation to single strand breaks of DNA. *Toxicol. Appl. Pharmacol* 148, 195-204.
7. Smith, C.J., Perfetti, T.A., Garg, R. and Hansch, C. (2003). IARC carcinogens reported in cigarette mainstream smoke and their calculated log P values. *Food Chem. Toxicol* 41, 807-817.
8. De Flora, S., D'Agostini, F., Balansky, R., Camoirano, A., Bennicelli, C., Bagnasco, M., Cartiglia, C., Tampa, E., Longobardi, M.G., Lubet, R.A. and Izzotti, A. (2003). Modulation of cigarette smoke-related end-points in mutagenesis and carcinogenesis. *Mutat. Res* 523-524, 237-252.
9. Van Schooten, F.J., Hirvonen, A., Maas, L.M., De Mol, B.A., Kleinjans, J.C., Bell, D.A. and Durrer, J.D. (1998). Putative susceptibility markers of coronary artery disease: association between VDR genotype, smoking, and aromatic DNA adduct levels in human right atrial tissue. *Faseb J* 12, 1409-1417.
10. Rockett, J.C., Burczynski, M.E., Fornace, A.J., Herrmann, P.C., Krawetz, S.A. and Dix, D.J. (2004). Surrogate tissue analysis: monitoring toxicant exposure and health status of inaccessible tissues through the analysis of accessible tissues and cells. *Toxicol. Appl. Pharmacol* 194, 189-199.
11. Godschalk, R.W., Moonen, E.J., Schilderman, P.A., Broekmans, W.M., Kleinjans, J.C. and Van Schooten, F.J. (2000). Exposure-route-dependent DNA adduct formation by polycyclic aromatic hydrocarbons. *Carcinogenesis* 21, 87-92.
12. 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.
13. Lampe, J.W., Stepaniants, S.B., Mao, M., Radich, J.P., Dai, H., Linsley, P.S., Friend, S.H. and Potter, J.D. (2004). Signatures of environmental exposures using peripheral leukocyte gene expression: tobacco smoke. *Cancer Epidemiol. Biomarkers Prev* 13, 445-453.
14. Hu, T., Gibson, D.P., Carr, G.J., Torontali, S.M., Tiesman, J.P., Chaney, J.G. and Aardema, M.J. (2004). Identification of a gene expression profile that discriminates indirect-acting genotoxins from direct-acting genotoxins. *Mutat Res* 549, 5-27.
15. Amundson, S.A., Do, K.T., Shahab, S., Bittner, M., Meltzer, P., Trent, J. and Fornace, A.J., Jr. (2000). Identification of potential mRNA biomarkers in peripheral blood lymphocytes for human exposure to ionizing radiation. *Radiat. Res* 154, 342-346.
16. Moore, M.M., Honma, M., Clements, J., Harrington-Brock, K., Awogi, T., Bolcsfoldi, G., Cifone, M., Collard, D., Fellows, M., Flanders, K., Gollapudi, B., Jenkinson, P., Kirby, P., Kirchner, S., Kraycer, J., McEnaney, S., Muster, W., Myhr, B., O'Donovan, M., Oliver, J., Ouldelhkim, M.C., Pant, K., Preston, R., Riach, C., San, R., Shimada, H. and Stankowski, L.F., Jr. (2002). Mouse lymphoma thymidine kinase gene mutation assay: follow-up International Workshop on Genotoxicity Test Procedures, New Orleans, Louisiana, April 2000. *Environ. Mol. Mutagen* 40, 292-299.
17. Kirsch-Volders, M., Sofuni, T., Aardema, M., Albertini, S., Eastmond, D., Fenech, M., Ishidate, M., Jr., Kirchner, S., Lorge, E., Morita, T., Norppa, H., Surrallés, J., Vanhauwaert, A. and Wakata, A. (2003). Report from the in vitro micronucleus assay working group. *Mutat. Res* 540, 153-163.
18. Rustemeier, K., Stabbert, R., Haussmann, H.J., Roemer, E. and Carmines, E.L. (2002). Evaluation of the potential effects of ingredients added to cigarettes. Part 2: chemical composition of mainstream smoke. *Food Chem. Toxicol* 40, 93-104.

19. Eller, P.M. NIOSH Manual of Analytical Methods (Vol. 2), National Institute of Occupational Safety and Health, USA, Cincinnati, OH, USA, 1984.
20. NIOSH Polynuclear aromatic hydrocarbons by HPLC (issue 3), The National Institute for Occupational Safety and Health, USA, 1998.
21. Gonzalez Borroto, J.I., Creus, A. and Marcos, R. (2001). Genotoxic evaluation of the furylethylene derivative 2-furyl-1-nitroethene in cultured human lymphocytes. *Mutat. Res* 497, 177-184.
22. Hasseman, J. Standard Operating Procedure: Aminoallyl labeling of RNA for microarrays (eukaryotic), The Institute for Genomic Research. Available: http://pga.tigr.org/sop/M004_1a.pdf [effective date 3/4/02], 2002.
23. Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4, P3.
24. Hosack, D.A., Dennis, G., Jr., Sherman, B.T., Lane, H.C. and Lempicki, R.A. (2003). Identifying biological themes within lists of genes with EASE. *Genome Biol* 4, R70.
25. Reddy, M.V. and Randerath, K. (1986). Nuclease P1-mediated enhancement of sensitivity of ³²P-postlabeling test for structurally diverse DNA adducts. *Carcinogenesis* 7, 1543-1551.
26. Godschalk, R.W., Maas, L.M., Van Zandwijk, N., van 't Veer, L.J., Breedijk, A., Borm, P.J., Verhaert, J., Kleinjans, J.C. and van Schooten, F.J. (1998). Differences in aromatic-DNA adduct levels between alveolar macrophages and subpopulations of white blood cells from smokers. *Carcinogenesis* 19, 819-825.
27. Nia, A.B., Maas, L.M., Brouwer, E.M., Kleinjans, J.C. and Van Schooten, F.J. (2000). Comparison between smoking-related DNA adduct analysis in induced sputum and peripheral blood lymphocytes. *Carcinogenesis* 21, 1335-1340.
28. Dertinger, S.D., Nazarenko, D.A., Silverstone, A.E. and Gasiewicz, T.A. (2001). Aryl hydrocarbon receptor signaling plays a significant role in mediating benzo[a]pyrene- and cigarette smoke condensate-induced cytogenetic damage in vivo. *Carcinogenesis* 22, 171-177.
29. Landi, M.T., Bertazzi, P.A., Baccarelli, A., Consonni, D., Masten, S., Lucier, G., Mocarelli, P., Needham, L., Caporaso, N. and Grassman, J. (2003). TCDD-mediated alterations in the AhR-dependent pathway in Seveso, Italy, 20 years after the accident. *Carcinogenesis* 24, 673-680.
30. van Delft, J.H., van Aken, E., van Breda, S.G., Herwijnen, M.H., Staal, Y.C. and Kleinjans, J.C. (2004). Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling. *Carcinogenesis* 25, 1265-1276.
31. Nordskog, B.K., Blixt, A.D., Morgan, W.T., Fields, W.R. and Hellmann, G.M. (2003). Matrix-degrading and pro-inflammatory changes in human vascular endothelial cells exposed to cigarette smoke condensate. *Cardiovasc. Toxicol* 3, 101-117.
32. van Grevenynghe, J., Sparfel, L., Le Vee, M., Gilot, D., Drenou, B., Fauchet, R. and Fardel, O. (2004). Cytochrome P450-dependent toxicity of environmental polycyclic aromatic hydrocarbons towards human macrophages. *Biochem. Biophys. Res. Commun* 317, 708-716.
33. De Jong, W.H., Kroese, E.D., Vos, J.G. and Van Loveren, H. (1999). Detection of immunotoxicity of benzo[a]pyrene in a subacute toxicity study after oral exposure in rats. *Toxicol Sci* 50, 214-220.
34. Ryder, M.I., Hyun, W., Loomer, P. and Haqq, C. (2004). Alteration of gene expression profiles of peripheral mononuclear blood cells by tobacco smoke: implications for periodontal diseases. *Oral Microbiol. Immunol* 19, 39-49.
35. Amundson, S.A., Grace, M.B., McLeland, C.B., Epperly, M.W., Yeager, A., Zhan, Q., Greenberger, J.S. and Fornace, A.J., Jr. (2004). Human in vivo radiation-induced biomarkers: gene expression changes in radiotherapy patients. *Cancer Res* 64, 6368-6371.
36. Hellermann, G.R., Nagy, S.B., Kong, X., Lockey, R.F. and Mohapatra, S.S. (2002). Mechanism of cigarette smoke condensate-induced acute inflammatory response in human bronchial epithelial cells. *Respir. Res* 3, 22.
37. Gohl, G., Lehmkoetter, T., Munzel, P.A., Schrenk, D., Viebahn, R. and Bock, K.W. (1996). TCDD-inducible plasminogen activator inhibitor type 2 (PAI-2) in human hepatocytes, HepG2 and monocytic U937 cells. *Carcinogenesis* 17, 443-449.
38. Fesus, L., Thomazy, V. and Falus, A. (1987). Induction and activation of tissue transglutaminase during programmed cell death. *FEBS Lett* 224, 104-108.

Chapter 3

Cigarette smoke-induced differential gene expression in blood cells from monozygotic twin pairs

D.M. van Leeuwen¹

E. van Agen¹

R.W.H. Gottschalk¹

R. Vlietinck²

M. Gielen²

M.H.M. van Herwijnen¹

L.M. Maas¹

J.C.S. Kleinjans¹

J.H.M. van Delft¹

¹ Department of Health Risk Analysis and Toxicology,
Maastricht University, Maastricht, the Netherlands

² Department of Population Genetics, Genomics and Bioinformatics,
Maastricht University, Maastricht, the Netherlands

Carcinogenesis (in press)

Abstract

Chemical carcinogenesis induced by life style factors like cigarette smoking is a major research area in molecular epidemiology. Gene expression analysis of large numbers of genes simultaneously using microarrays holds the opportunity to study the effects of such an exposure at the genome level yielding more mechanism-based information. Therefore, the aim of our study was to investigate multiple gene expressions in blood, indicative for the effects caused by cigarette smoke. Smoking-discordant monozygotic twin pairs (n=9) were studied to diminish influences of genetic background. Using a dedicated microarray containing 600 toxicologically relevant genes, we investigated which genes are differentially expressed in smokers compared to non-smokers. We also looked for genes of which the expression changes correlated with DNA-adducts, a bio-marker of effective dose for exposure to cigarette smoke carcinogens. The mean DNA adduct level in smokers differed significantly from that in non-smokers (2.0 ± 0.2 versus 1.2 ± 0.2 adducts/ 10^8 nucleotides; $p=0.04$). The genes for which the expression differed most significantly between smokers and non-smokers are *ATF4*, *MAPK14*, *SOD2*, *CYP1B1* and *SERPINB2*. *CYP1B1* and *SOD2* can directly be linked to cigarette smoke exposure, whereas the other genes are associated with stress or environmentally induced response. Main functions of the genes influenced by cigarette smoking comprise carcinogen metabolism, oxidative stress response and anti-apoptosis.

Introduction

Chemical carcinogenesis induced by life style factors like cigarette smoking is a major research area in molecular epidemiology. Cigarette smoke is a well known source of chemical carcinogens and comprises a complex mixture of over 60 proven, probable and possible carcinogenic agents [1]. It is known to induce a series of biomarkers for genotoxic, pre-carcinogenic effects, both in target tissue as in surrogate target tissue like peripheral blood mononuclear cells (PBMC) [2,3]. Many of these early events or effects, e.g. DNA adducts, micronuclei, chromosome aberrations and mutations, are detectable in blood cells long before health effects appear [4,5].

Gene expression profiling by DNA microarray technology is considered a promising tool to gain more knowledge on the effects of toxicant exposure on the transcriptome level [6-8]. The development of tools for analyzing differential gene expression of many genes simultaneously has provided the opportunity for analyzing effects in exposed populations at the genome level, thereby studying the biological response to environmental exposure in an integrative manner.

Differential gene expression in human peripheral blood cells *in vivo* has been reported in a few studies [9-12]. In a study on human leukocyte gene expression in smokers versus non-smokers, particular gene expression modulations have been found to correlate significantly with plasma cotinine levels and these genes accurately distinguished smokers from non-smokers [10]. No study has yet associated gene expression in exposed populations with established biomarkers, like DNA-adducts, which would provide an important opportunity for phenotypically anchoring gene expression profiles.

The goal of our study therefore, was to investigate which genes are differentially expressed in blood cells from smokers as compared to non-smokers and for which genes the changes in expression correlate with aromatic DNA-adduct levels. In order to reduce the potential impact from differences in genetic background, monozygotic twin pairs that were discordant in their smoking behavior were used as study subjects, implying that each twin pair consisted of one smoker and one non-smoker. Plasma levels of cotinine (a stable metabolite of nicotine) levels were measured to confirm smoking behavior. We used a toxicologically focused cDNA microarray to analyze gene expression differences as we anticipate that potential biomarker genes for cigarette smoke exposure should be represented within such a typical selection of target genes.

Materials and methods

Study population and blood collection

Participants were registered members of the East-Flanders Prospective Twin Survey (EFPTS) [13]. All twins over age of 20 years were sent questionnaires to collect information on smoking behavior. Twin pairs were included if they reported to be monozygotic and discordant in cigarette-smoking behavior (per twin pair one individual is a smoker whereas the other is a non-smoker). Initially, the study population consisted of 10 pairs of smoking-discordant twins; 5 male-male pairs and 5 female-female pairs. Genotyping for several biotransformation and DNA repair genes indicated that the two individuals of

one male-male twin pair differed in 6 out of 19 polymorphisms (data not shown) and thus appeared to be a dizygotic twin pair. This twin pair was therefore excluded from further analysis. Among the non-smokers, there were 7 never smokers. Two were former smokers and quit smoking at least 13 months prior to the study.

Venous blood was collected into one tube containing ethylenediamine-tetraacetic acid (EDTA) and five PAXgene Blood RNA tubes (PreAnalytix, Qiagen, Hilden, Germany). Tubes were kept at room temperature for at least 2 hours but maximally one day, and thereafter stored at 4°C until DNA and RNA isolation. Subjects gave informed consent and the study was approved by the Local Committee of Medical Ethics of the Catholic University of Leuven, Belgium.

Total RNA isolation

Total RNA was isolated within a week after blood collection according to the manufacturer's instructions with minor modifications using the PAXgene Blood RNA Kit (PreAnalytix, Qiagen, Hilden, Germany). The entire procedure was carried out at room temperature. Samples that did not show a distinct RNA-debris phase separation were submitted to an extra centrifugation step. RNA purity was determined by spectrophotometer, based on A260/A280 ratios. RNA integrity was determined on the BioAnalyzer and RNA was regarded intact when showing two distinct bands for 28S and 18S RNA.

Gene expression analyses by cDNA microarrays

A common reference RNA pool was made from aliquots of all individuals that yielded sufficient RNA. On all arrays, three RNA samples were hybridized together [14], *i.e.* the common reference sample and samples from the two individuals of a twin pair: the smoker and the non-smoker (for hybridization design, see Table 1). This allowed for analyzing differential gene expression at the level of twin pairs, and also at group levels for non-smokers and smokers. Total RNA was reverse transcribed and labeled for cDNA microarray analysis according to the TIGR protocol for amino-allyl labeling of total eukaryotic RNA for microarrays as described previously [15]. For each individual, three aliquots of total RNA were used and labeled with Cy5, Cy3 (Amersham Biosciences, Freiburg, Germany) or Alexa Fluor 594 (ARES™ DNA labeling kit, Molecular Probes, Leiden, the Netherlands), according to the manufacturers' instructions. For each twin pair, three hybridizations were performed, using a design in which each sample was labeled by each dye once (see Table 1). Equal amounts of labeled cDNA were combined and hybridized to Phase1 Human Tox 600 cDNA microarrays (Phase-1 Molecular Toxicology, Sante Fe, NM), which contains genes in quadruplicate, in toxicologically relevant categories such as cell cycle, proliferation, biotransformation, inflammation and apoptosis.

Table 1. Hybridization design for RNA samples from every twin pair using three fluorophores

Array	RNA sample		
	Smoker	Non-smoker	Reference
1	Cy5	Cy3	Alexa594
2	Alexa594	Cy5	Cy3
3	Cy3	Alexa594	Cy5

Microarrays were scanned on a ScanArrayExpress scanner (Perkin Elmer, Boston, MA) with a fixed laser power and a variable photomultiplier gain, such that no spots were saturated. TIFF images for the separate fluorophores were imported into Imagen (BioDiscovery, Marina del Rey, CA) and translated into raw data files, which were transformed and analyzed using GeneSight (BioDiscovery). Local background was subtracted and low expression values (< 5 pixel intensities) and bad spots were omitted. Thereafter, data were log₂ transformed and normalized with Locally Weighed Scatterplot Smoothing (LOWESS). After calculation of the differences, replicate spots were combined omitting outliers greater than 2 times the standard deviation, so that eventually the mean intensities were based on 12 replicates/spot (3 arrays x 4 replicates per array).

Gene expression analyses by real-time quantitative PCR

Real-time quantitative PCR (qPCR) was performed to validate the microarray results for a selection of genes. cDNA was generated from 2 µg of total RNA according to the manufacturer's instructions using the Bio-Rad iScript cDNA synthesis kit (Bio-Rad Laboratories, Hercules, CA). Aliquots were used for real-time qPCR on the BioRad MyiQ iCycler Single Color real-time PCR detection system using iQ™ SYBR® Green Supermix, containing all ingredients; iTaq Polymerase, dNTPs, SYBR Green I and buffers. Reactions were initiated for 3 min at 95°C, then 40 cycles of 15 seconds at 95°C and 45 seconds at 60°C were run. Melting curve analysis was performed starting at 60°C with step-wise temperature elevations of 0.5°C every 10 seconds. Beta-actin (*ACTB*), cyclophilin A (*PPIA*) and beta-2-microglobulin (*B2M*) were included as reference genes (internal controls) and all reactions were performed in duplicate. No template controls (water controls) and a dilution series of a pool of all cDNA samples from the study were included to estimate PCR efficiency. Primer sequences used were as follows:

ACTB

5'-CTGGGCACCCAGCACAAT-3' (forward) and 5'-GCCGATCCACACGGAGTACT-3' (reverse),

B2M

5'-TGACTTTGTACAGCCCAAGATA-3' (forward) and 5'-AATGCGGCATCTTCAAACCT-3' (reverse),

PPIA

5'-TTCCTGCTTTCACAGAATTATTCC-3' (forward) and 5'-GCCACCACTGCCATTATGG-3' (reverse),

ATF4

5'-CTCCAGCGACAAGGCTAAGG-3' (forward) and 5'-GTTGTTGGAGGGACTGACCAA-3' (reverse),

MAPK14

5'-TGAAGACTGTGAGCTGAAGATTCTG-3' and 5'-CCACGTAGCCTGTCAATTCATC-3' (reverse),

SOD2

5'-ATCAGGATCCACTGCAAGGAA-3' (forward) and 5'-CGTGCTCCCACACATCAATC-3' (reverse),

PTEN

5'-CAGTGGCGGAAGTCTGCAAT-3' (forward) and 5'-CGTCGTGTGGGTCTGAATT-3' (reverse),

APP

5'-CCGCTCTGCAGGCTGTTC-3' (forward) and 5'-CGCGGACATACTTCTTTAGCATATT-3' (reverse),

AXIN1

5'-GAGAGCCCCAAAGTCTGTAGTGA-3' (forward); 5'-AAGGTCGGCAGGTATCCAGAT-3' (reverse)

BAK1

5'-GAATGCCTATGAGTACTTCACCAAGA-3' (forward); 5'-ACACGGCCCCAATTGATG-3' (reverse)

CSF1R

5'-CAGCACCAACAACGCTACCTT-3' (forward) and 5'-CGGGCAGGGTCTTTGACATA-3' (reverse),

IL-10

5'-GGCGCTGTCATCGATTCTT-3' (forward); 5'-TGGAGCTTATTAAGGCATTCTTCA-3' (reverse),

OGG1

5'-AGAGGTGGCTCAGAAATCCAA-3' (forward); 5'-CAGATAAAGAGAAAAGGCATTCTGA-3' (reverse),

PCK2

5'-CAAGACCAACCTGGCTATGATG-3' (forward); 5'-GAGTCGACCTTCACTGTCAAACC-3' (reverse),

CYP1B1

5'-AGTGCAGGCAGAAATTGGATCA-3' (forward); 5'-GCGCATGGCTTCATAAAGGA-3' (reverse) and

SERPINB2

5'-CGATTTTGCAGGCACAAGCT-3' (forward) and 5'-CCTGTGGATGCATTGATTGC-3' (reverse).

Data analysis

Computational and statistical analyses of the microarray data were performed using software programs GeneSight (BioDiscovery) and SPSS 11.5 (SPSS Inc., Chicago, IL). Gene expressions as modulated by cigarette smoking, were identified using several approaches; either by analyzing twin pairs (the expression differences between a smoker and non-smoker directly), or by analyzing each person individually (the expression difference between an individual and the common reference) through pair wise analyses. Confidence analysis, performed in GeneSight at >99% confidence level and >0.1 regulation level, was used to find differentially expressed genes within the twin pairs. For testing differences between smokers and non-smokers, using the expression differences between each individual and the common reference, the non-parametric 2-related samples Wilcoxon test in SPSS 11.5 was used. Next to these two tests, the signal-to-noise ratio for the smokers versus the non-smokers was calculated:

$$(\text{mean difference}_{\text{smokers/ref}} - \text{mean difference}_{\text{non-smokers/ref}}) / (\text{stdev difference}_{\text{smokers/ref}} + \text{stdev difference}_{\text{non-smokers/ref}})$$

The combination of these approaches was used in order to limit the number of false-positives. In addition, Spearman correlation tests were performed to investigate significant correlations of individual gene expressions with individual aromatic DNA-adduct levels.

For the real-time qPCR analyses, ΔCt or $\Delta\Delta\text{Ct}$ values were calculated for each gene of interest [16]. Three different approaches for normalizing these data were used: 1) β -actin was used as the housekeeping or 'stably expressed' gene, 2) the average of three housekeeping or 'stably expressed' genes was used (beta-actin *ACTB*, beta-2-microglobulin *B2M* and cyclophilin A *PPIA*), or 3) no house keeping genes were used but the amount of input RNA for a reference and ΔCt values were calculated directly for the gene of interest by subtracting the common reference Ct from the test sample Ct.

Gene ontologies were retrieved from <http://www.ncbi.nlm.nih.gov> with LocusLink or Gene. Analysis of biologically enriched pathways was performed in the Expression Analysis Systematic Explorer (EASE, <http://david.niaid.nih.gov/ease>) [17].

Cotinine measurements

Plasma was collected from EDTA blood by centrifugation. Plasma cotinine was measured by high performance liquid chromatography (HPLC) according to the protocol described by Van Vunakis *et al.* [18] with the exception of the presence of EDTA in the standard. The significance of the difference between smokers and non-smokers was analyzed with a Wilcoxon test for non-parametric analysis of 2 related samples in SPSS 11.5.

DNA adduct measurements

Leukocyte fractions were isolated from whole blood (after removal of the plasma layer) through erythrocyte lysis (lysis buffer containing 155 mM NH₄Cl, 10 mM KHCO₃, 10 mM EDTA :blood = 3:1) for 30 min at 4°C. Leukocytes were stored at -80°C until DNA isolation. DNA was isolated by standard phenol extraction [19]. DNA adduct levels were measured by ³²P-postlabeling with Nuclease-P1 enrichment as originally described by Reddy and Randerath [20] with minor modifications as described by Godschalk *et al.* [19]. The significance of the difference between smokers and non-smokers was analyzed with a Wilcoxon non-parametric test for 2 related samples in SPSS 11.5.

Results

Differential gene expression in blood caused by cigarette smoking was studied in blood cells from monozygotic twins discordant in smoking behavior. The characteristics of the study population are shown in Table 2. The average plasma cotinine level in non-smokers was significantly lower than in smokers, namely 5.6 ± 0.2 ng/ml versus 247.6 ± 54.7 ng/ml ($p=0.008$). Average DNA adduct levels were 1.2 ± 0.2 adducts/ 10^8 nucleotides in the non-smokers group; significantly different ($p=0.04$) from the average of 2.0 ± 0.2 adducts/ 10^8 nucleotides in the smokers group.

Table 2. Study population characteristics

Number of twins	9
Age	27.5 ± 1.7
Cigarettes/day	14.2 ± 2.9 17.2 ± 2.0**
Pack years	7.5 ± 1.8
Cotinine in plasma (ng/ml)	
smokers	247.6 ± 54.7
non-smokers	5.6 ± 0.2
difference	p = 0.008
DNA adducts /10 ⁸ normal nucleotides	
smokers	1.95 ± 0.21 (n=9)
non-smokers	1.17 ± 0.16 (n=7)***
difference	p = 0.04

All values are presented as mean ± standard error of the mean

* Lowest number of cigarettes/day reported

** Highest number of cigarettes/day reported

*** For two non-smokers DNA adduct analyses could not be performed.

For differential gene expression analysis, three RNA samples were hybridized on every cDNA microarray, namely of the smoker and non-smoker samples from the same twin pair and a common reference sample. Through a combination of several approaches genes were identified with altered expression due to cigarette smoking. Either data of twin couples (the expression differences between a smoker and non-smoker), or data from each person individually (the expression difference between an individual and the common reference) were analyzed. Using the data of twin pairs, confidence analysis revealed 34 differentially expressed genes. Using the data from smoking and non-smoking individuals, 76 differentially expressed genes in smokers compared to non-smokers were revealed by a non-parametric Wilcoxon signed ranks test for 2 related samples. In addition, 44 genes showed a signal-to-noise ratio for the smokers versus the non-smokers of >0.5. The data of these tests are summarized in a VENN-diagram (Figure 1). The seven genes in the intersection of the diagram were identified by all three tests and appeared up regulated in the smokers.

Figure 1. VENN diagram representing genes differentially expressed due to smoking as revealed by various statistical approaches. Tests included are the Wilcoxon non-parametric signed ranks test for 2 paired samples, confidence analysis (confidence level 99%, and expression difference >0.1 on a log2 scale) and signal-to-noise ratio test. The seven genes in the inter-section were revealed by all 3 approaches.

Second, genes that correlated with DNA adducts were identified by Spearman correlation tests. Seven genes that scored high on correlation coefficient are reported in Table 3. For most genes, such as *OGG1*, gene expression changes appeared to be negatively correlated with DNA adduct levels.

Table 3. Spearman correlations of gene expression differences of each individual relative to the common reference with DNA adduct levels

	correlation (R)	p-value
<i>BAK1</i>	-0.54	0.029
<i>CSF1R</i>	-0.51	0.043
<i>IL-10</i>	-0.56	0.024
<i>APP</i>	0.70	0.003
<i>AXIN1</i>	-0.54	0.029
<i>OGG1</i>	-0.69	0.003
<i>PCK2</i>	-0.57	0.022

For a selection of genes that might be most interesting for future molecular epidemiology studies, the expression differences were verified by means of real-time qPCR. For the cigarette smoking related effects, *SOD2*, *PTEN*, *ATF4* and *MAPK14* were selected. *OGG1*, *IL10*, *APP*, *AXIN1*, *PCK2*, *BAK* and *CSF1R* were selected based on correlations with DNA adduct levels. Furthermore, plasminogen activator inhibitor 2 (*SERPINB2*) was included based on significant differential expression in previous *in vitro* studies with pe-

ripheral blood mononuclear cells [24]. Cytochrome P450 1B1 (*CYP1B1*) was included as this gene is known to be affected by *e.g.* PAHs found in cigarette smoke and to be up-regulated in smokers [21-23] and it also responded strongly to exposure to cigarette smoke condensate *in vitro* (unpublished data).

For *SOD2*, *MAPK14* and *ATF4*, the microarray data were reproduced by qPCR, especially when applying the ΔCt method (Table 4). For *PTEN*, qPCR results showed no significant difference between smokers and non-smokers, in contrast to the microarray data. For *SERPINB2* and *CYP1B1* however, qPCR showed differential expression in smokers versus non-smokers (p-values 0.021 and 0.008 respectively), where microarray analysis did not report an effect. When qPCR gene expressions for *OGG1*, *IL10*, *APP*, *AXIN1*, *PCK2*, *BAK* and *CSF1R* were correlated with DNA adduct levels, the correlations found for the microarray data with DNA adducts could not be reproduced (data not shown). Ontologies of these genes are shown in Table 5.

Table 4. Comparison of differential gene expressions between smokers and non-smokers as measured by two independent assays: cDNA microarrays and real-time PCR

Gene	Assay	T-test (Wilcoxon p- value)	Mean differ- ence smoker- non smoker (log2)	Standard error of the mean
<i>SOD2</i>	microarray	0.011	0.26	0.09
	$\Delta\Delta\text{Ct}$ (<i>ACTB</i>) [*]	0.038	0.27	0.11
	$\Delta\Delta\text{Ct}$ (<i>ACTB+PPIA+B2M</i>) ^{**}	0.109	0.34	0.15
	ΔCt ^{***}	0.011	0.53	0.14
<i>MAPK14</i>	microarray	0.008	0.20	0.08
	$\Delta\Delta\text{Ct}$ (<i>ACTB</i>)	0.123	0.26	0.16
	$\Delta\Delta\text{Ct}$ (<i>ACTB+PPIA+B2M</i>)	0.213	0.33	0.22
	ΔCt	0.021	0.54	0.25
<i>ATF4</i>	microarray	0.011	0.26	0.09
	$\Delta\Delta\text{Ct}$ (<i>ACTB</i>)	0.374	-0.12	0.11
	$\Delta\Delta\text{Ct}$ (<i>ACTB+PPIA+B2M</i>)	0.440	-0.05	0.07
	ΔCt	0.050	0.15	0.06
<i>PTEN</i>	microarray	0.038	0.17	0.07
	$\Delta\Delta\text{Ct}$ (<i>ACTB</i>)	0.343	0.09	0.16
	$\Delta\Delta\text{Ct}$ (<i>ACTB+PPIA+B2M</i>)	0.374	0.15	0.19
	ΔCt	0.139	0.35	0.23
<i>CYP1B1</i>	microarray	0.575	0.10	0.24
	$\Delta\Delta\text{Ct}$ (<i>ACTB</i>)	0.008	0.59	0.15
	$\Delta\Delta\text{Ct}$ (<i>ACTB+PPIA+B2M</i>)	0.008	0.66	0.19
	ΔCt	0.008	0.86	0.24
<i>SERPINB2</i>	microarray	0.233	-0.09	0.07
	$\Delta\Delta\text{Ct}$ (<i>ACTB</i>)	0.066	0.38	0.17
	$\Delta\Delta\text{Ct}$ (<i>ACTB+PPIA+B2M</i>)	0.038	0.44	0.18
	ΔCt	0.015	0.64	(0.20)

* β -actin (*ACTB*) was used as the housekeeping or 'stably expressed' gene

** the average of three housekeeping or 'stably expressed' genes was used (*ACTB*, *B2M* and *PPIA*)

*** no house keeping genes were used but the amount of input RNA for normalization.

Table 5. Gene Ontologies of differentially expressed genes

GenBank accession #	Abbreviation and gene name	Effect	Molecular function / biological process*	Main function*
D90209	<i>ATF4</i> Activating transcription factor 4	↑	DNA binding; RNA polymerase II transcription factor activity; regulation of transcription, DNA-dependent	Transcription factor
L35253	<i>MAPK14</i> Mitogen-activated protein kinase 14	↑	ATP binding; MAP kinase activity; cell motility cell surface receptor linked signal transduction; chemotaxis; protein amino acid phosphorylation; protein serine/threonine kinase activity; response to stress; transferase activity	MAPK family; involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development
XM_005867	<i>PTEN</i> Phosphatase and tensin homolog (mutated in multiple advanced cancers 1)	↑	Hydrolase activity; phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase activity; protein tyrosine phosphatase activity; negative regulation of cell cycle; protein amino acid dephosphorylation	Tumor suppressor by negative regulation AKT/PKB signaling pathway
Y00985	<i>SOD2</i> Superoxide dismutase 2, mitochondrial	↑	Manganese ion binding; manganese superoxide dismutase activity; metal ion binding; oxidoreductase activity; response to oxidative stress; superoxide metabolism	Free radical scavenging enzyme
L08246	<i>MCL1</i> Myeloid cell leukemia sequence 1 (BCL2-related)	↑	Anti-apoptosis; cell differentiation; cell fate determination; cell homeostasis; protein binding; protein channel activity; protein heterodimerization activity	Regulation of apoptosis
K01396	<i>SERPINA1</i> Serine (or cysteine) proteinase inhibitor, clade A (alpha-1, antitrypsin), member 1	↑	Acute-phase response; protein binding; serine-type endopeptidase inhibitor activity	Protease inhibitor, deficiency is associated with emphysema and liver disease
U14588	<i>PAX</i> Paxillin	↑	Cell motility; cell-matrix adhesion; cytoskeleton; microtubule associated complex; protein binding; signal complex formation; signal transduction	-
U03688	<i>CYP1B1</i> Cytochrome P450, family 1, subfamily B, polypeptide 1	↑	Electron transport; eye morphogenesis (sensu Mammalia); monooxygenase activity; oxygen binding; visual perception	Phase I metabolism of procarcinogens such as PAHs
M18082	<i>SERPINB2</i> Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 2	↑	Anti-apoptosis; plasminogen activator activity; serine-type endopeptidase inhibitor activity	Primary regulator of plasminogen activation

*According to Gene Ontologies from <http://www.ncbi.nlm.nih.gov/Gene>

Discussion

In nine monozygotic twin pairs discordant for smoking, we investigated whether cigarette smoking causes differential gene expression of toxicologically relevant genes in peripheral blood cells and whether expression levels correlate with a biomarker for early effects, *i.e.* DNA adducts. By analyzing cigarette smoke-induced differential gene expression in monozygotic twins, we reduced the impact of inter-individual variability due to variation in genetic background. It also provided the opportunity to perform pair-wise analyses, adding statistical power to the study. The analyses revealed several genes to be reproducibly differentially expressed due to cigarette smoking.

Genes which were differentially expressed in smokers compared to non-smokers were identified by a combination of several approaches, *viz.* non-parametric Wilcoxon test for 2 related samples, confidence analysis and signal-to-noise calculation. We choose for this combination of different approaches as it is likely to reduce the number of false-positive calls. The genes that were found by all three approaches (Figure 1) were considered the most discriminatively relevant genes. Resulting genes are activating transcription factor 4 (*ATF4*), mitogen-activated protein kinase 14 (*MAPK14* aka *p38*), superoxide dismutase 2/Mn (*SOD2*), alpha 1-antitrypsin (*SERPINA1*), paxillin (*PXN*), myeloid cell differentiation protein-1 (*MCL1*) and phosphate and tensin homolog (*PTEN*). These genes were all up-regulated in smokers as compared to non-smokers.

ATF4, *SOD2* and *MAPK14* were the genes differentially expressed by cigarette smoking, which were confirmed by real-time PCR. In addition to these genes, two other genes that did not show significance in the microarray results were included in the real-time PCR analyses, *i.e.* *CYP1B1* and *SERPINB2*. In the real-time PCR analysis, *CYP1B1* showed, for every twin pair, a higher expression in the smoker than in the non-smoker, which agrees with other studies [21-23]. However, microarray analysis did not demonstrate a significant effect for *CYP1B1* due to the relatively high standard error of the mean, which is probably due to the low mRNA abundance of this gene in peripheral blood cells. Also, *SERPINB2* showed significance in the real-time PCR analysis, which agrees with previous in vitro experiments on human lymphocytes treated with cigarette smoke condensate [24]. Upon microarray analysis, no significant effects were observed for this gene, again possibly due to the low abundance of its mRNA.

Based on the current data, the genes that we would propose to be suitable for biomonitoring in cigarette smoke-exposed humans are *ATF4*, *SOD2*, *MAPK14*, *CYP1B1* and *SERPINB2* since they showed to be reproducibly differentially expressed as a result of exposure to cigarette smoke. The ontologies for these genes are shown in Table 5. *CYP1B1* is a well known phase-1 biotransformation enzyme involved in the metabolism of carcinogens. Our data confirm previous studies, where increased *CYP1B1* mRNA levels have already been found in smokers and lung cancer patients [21-23]. Whether expression changes for this gene reflect exposure rather than genotoxic effect, is not clear from the current data. *SOD2* is known to be involved in (response to) oxidative stress and superoxide metabolism, scavenging intracellular oxidants, thus functioning as a defense mechanism. *SOD2* has been found to be triggered by smoking as it has been found elevated in the lungs of smokers [25]. *MAPK14* has not directly been linked to exposure to cigarette smoke, but more generally to environmental exposures and stress conditions [26,27]. *MAPK14* signaling is suggested to play a crucial role in the regulation

of cellular proliferation and differentiation and recent studies suggest a (main) role in protecting cells from apoptosis [28,29]. Therefore this gene might be hypothesized to mark an effect rather than an exposure. *ATF4* activation has been reported in HepG2 cells after arsenite exposure [30], but no data are available for primary (human) cells or tissues from exposed populations. For *SERPINB2*, a role has been suggested in (malignancy of) endometrial cancer [31] and this gene is expressed in many tumors [32]. The gene has also shown to confer resistance to apoptosis [33]. Members of the *SERPIN* gene family of serine protease inhibitors have also been reported to be involved in COPD, emphysema and associated with lung cancer risk [34-36]. Induction of differential gene expression of these five genes by cigarette smoking could therefore be the subject of further investigation in larger populations.

Our second approach was to identify genes for which the expression is associated with an established biomarker for early effect, like DNA-adducts (see Table 3). Although DNA adduct levels significantly differed between smokers and non-smokers (p-value 0.04), a clear overlap existed between both groups. Therefore, gene expressions associated with DNA adduct levels, not necessarily have to discriminate smokers from non-smokers. Indeed, none of the genes differentially expressed between smokers and non-smokers were in the list of genes correlating with DNA adducts. Another explanation for lack of overlap between both gene lists might be that cigarette smoke is composed of much more carcinogens and chemicals than those that cause DNA adducts as measurable by ³²P-postlabelling. The adduct levels found in the study population, are well within range of values found in other studies [37]. Unfortunately, when verifying the expression of the correlating genes by real-time PCR and correlating these expression data with DNA adducts, no significances were found. Possibly this can be attributed to the small expression range of these genes, which in general is much smaller than that for the genes differentially expressed between smokers and non-smokers. This finding suggests lower specificity of DNA adducts as compared to differential gene expression for the measurement of effects of cigarette smoke exposure.

The genes we found to be differentially expressed in peripheral blood from cigarette smokers compared to non-smokers within monozygotic twin pairs, have not been reported by other, similar studies in which differential gene expression has been measured in vivo in peripheral blood [9-11]. In a study in leukocytes by Lampe *et al.*, differential gene expression in smokers and non-smokers was investigated and genes have been reported that correlated significantly with plasma cotinine levels [10]. For comparison with this study, we investigated the correlation of *CYP1B1* and *IL-1 β* (the only two genes reported by Lampe *et al.* that were also present on our microarrays) between the individual gene expression and individual cotinine levels. We did not find significant correlations of these genes with plasma cotinine levels in our data (Spearman test, $p > 0.05$). Reasons for this might be that Lampe *et al.* used a different platform and approach for selecting gene expressions and that they investigated a study population with a larger variation in subjects' age. Another recent study has investigated differential gene expression in blood from individuals exposed to metal fumes [11]. Upon analyzing the subgroup of smokers in that study, it was suggested that cigarette smoking considerably affected whole-blood gene expression profiles. It was also proven that, with a paired-samples study design, it was possible to measure small changes in gene expression in a population study, which is in line with our findings. However, also Wang *et al.* used another platform and did not verify their main findings with an independent tech-

nique like real-time PCR. A study in benzene-exposed factory workers reported to have found a set of genes as potential biomarkers of benzene exposure [12]. Although benzene is a component of cigarette smoke, it does not appear that after exposure to cigarette smoke benzene is an important component to alter gene expression, as no genes were reported that resemble the data of the present study. Again this study used another platform and a smaller study population.

Besides the above mentioned human studies, research on cigarette smoke-induced gene expression changes in animal models has been reported. In SKH-1 mice, cigarette smoke exposure induces *MAPK* expression [38,39], and expression of a *SERPIN* (trypsin-chemotrypsin related serine protease) gene [38]. In fetal liver of Swiss albino mice, exposure to environmental cigarette smoke induced expression of *CYP* genes as well as precursors for two *SODs* [39]. Several *MAPKs* have been shown to be up-regulated by cigarette smoke in the lungs of A/J mice [40].

When performing real-time PCR analyses of gene expression, the data can be normalized by either using the same amount of input RNA or by including housekeeping or stably expressed genes in the analyses [16]. When verifying our microarray data by real-time PCR for the differentially expressed genes, consistency between both methods depended on the approach to normalize the real-time PCR data. The best reproducibility of the microarray data was achieved by the ΔC_t method (see Table 4), where the amount of input RNA was used for normalization, instead of one or more house keeping genes.

The present study is one of the few studies that apply transcriptomics within the field of molecular epidemiology. Because of the limited history, the application of toxicogenomic tools in molecular epidemiology may be influenced by some limitations and pitfalls. For example, in toxicogenomic research microarrays are most often used in qualitative manner; to get an overview of expression changes of large numbers of genes. It is common practice to verify (the most relevant) findings of microarray experiments with real-time PCR, which represents a more quantitative assay [9,12,24]. As is discussed above, the results of both techniques can deviate. In relation to this, is the issue of (choosing the right) 'housekeeping' gene(s) to be included in the real-time PCR analyses for normalization. Another pitfall or limitation of microarray methodologies can be the issue of multi-comparison testing, which is prone to deliver false-positive results [12,41]. This potential problem was dealt with in the current study by performing different statistical analyses and regarding as the most interesting set of genes, only those that showed significant results in all of the tests. Further developments in the field of toxicogenomics and transcriptomics, the technologies and data analyses, will provide ways to handle these limitations and pitfalls in the future.

Summarizing, this study on cigarette smoking-induced differential gene expression in peripheral blood from monozygotic smoking-discordant twins, showed that gene expression analysis of many genes simultaneously yields a gene expression profile as promising biomarker of effect induced by cigarette smoking. Five genes were revealed that we would propose to be suitable for biomonitoring in a human population, namely *SOD2*, *MAPK14*, *ATF4*, *CYP1B1* and *SERPINB2*.

Acknowledgements

The authors thank Dr. C. Derom and the nurses of the East-Flanders Prospective Twin Survey (EFPTS) for blood collections and assistance on mailing the questionnaires. This study was financially supported by the Flanders Environment and Health Study.

References

1. Smith, C.J., Perfetti, T.A., Garg, R. and Hansch, C. (2003) IARC carcinogens reported in cigarette main-stream smoke and their calculated log P values. *Food Chem. Toxicol*, 41, 807-17.
2. Van Schooten, F.J., Hirvonen, A., Maas, L.M., De Mol, B.A., Kleinjans, J.C., Bell, D.A. and Durrer, J.D. (1998) Putative susceptibility markers of coronary artery disease: association between VDR genotype, smoking, and aromatic DNA adduct levels in human right atrial tissue. *Faseb J*, 12, 1409-17.
3. De Flora, S., D'Agostini, F., Balansky, R., Camoirano, A., Bennicelli, C., Bagnasco, M., Cartiglia, C., Tampa, E., Longobardi, M.G., Lubet, R.A. and Izzotti, A. (2003) Modulation of cigarette smoke-related end-points in mutagenesis and carcinogenesis. *Mutat. Res*, 523-524, 237-52.
4. van Delft, J.H., Baan, R.A. and Roza, L. (1998) Biological effect markers for exposure to carcinogenic compound and their relevance for risk assessment. *Crit. Rev. Toxicol*, 28, 477-510.
5. Bonassi, S. and Au, W.W. (2002) Biomarkers in molecular epidemiology studies for health risk prediction. *Mutat. Res*, 511, 73-86.
6. Hamadeh, H.K., Amin, R.P., Paules, R.S. and Afshari, C.A. (2002) An overview of toxicogenomics. *Curr Issues Mol Biol*, 4, 45-56.
7. Toraason, M., Albertini, R., Bayard, S., Bigbee, W., Blair, A., Boffetta, P., Bonassi, S., Chanock, S., Christani, D., Eastmond, D., Hanash, S., Henry, C., Kadlubar, F., Mirer, F., Nebert, D., Rapport, S., Rest, K., Rothman, N., Ruder, A., Savage, R., Schulte, P., Siemiatycki, J., Shields, P., Smith, M., Tolbert, P., Vermeulen, R., Vineis, P., Wacholder, S., Ward, E., Waters, M. and Weston, A. (2004) Applying new biotechnologies to the study of occupational cancer--a workshop summary. *Environ. Health Perspect*, 112, 413-6.
8. Aardema, M.J. and MacGregor, J.T. (2002) Toxicology and genetic toxicology in the new era of "toxicogenomics": impact of "-omics" technologies. *Mutat Res*, 499, 13-25.
9. 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, 1429-38.
10. Lampe, J.W., Stepaniants, S.B., Mao, M., Radich, J.P., Dai, H., Linsley, P.S., Friend, S.H. and Potter, J.D. (2004) Signatures of environmental exposures using peripheral leukocyte gene expression: tobacco smoke. *Cancer Epidemiol. Biomarkers Prev*, 13, 445-53.
11. Wang, Z., Neuburg, D., Li, C., Su, L., Kim, J.Y., Chen, J.C. and Christiani, D.C. (2005) Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. *Environ Health Perspect*, 113, 233-41.
12. Forrest, M.S., Lan, Q., Hubbard, A.E., Zhang, L., Vermeulen, R., Zhao, X., Li, G., Wu, Y.Y., Shen, M., Yin, S., Chanock, S.J., Rothman, N. and Smith, M.T. (2005) Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. *Environ Health Perspect*, 113, 801-7.
13. Derom, C., Vlietinck, R., Thiery, E., Leroy, F., Fryns, J.P. and Derom, R. (2002) The East Flanders Prospective Twin Survey (EFPTS). *Twin Res*, 5, 337-41.
14. Staal, Y.C., van Herwijnen, M.H., van Schooten, F.J. and van Delft, J.H. (2005) Application of four dyes in gene expression analyses by microarrays. *BMC Genomics*, 6, 101.
15. Hasseman, J. (2002) Standard Operating Procedure: Aminoallyl labeling of RNA for microarrays (eukaryotic). The Institute for Genomic Research. Available: http://pga.tigr.org/sop/M004_1a.pdf [effective date 3/4/02].
16. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods*, 25, 402-8.
17. Hosack, D.A., Dennis, G., Jr., Sherman, B.T., Lane, H.C. and Lempicki, R.A. (2003) Identifying biological themes within lists of genes with EASE. *Genome Biol*, 4, R70.
18. Van Vunakis, H., Gijka, H.B. and Langone, J.J. (1993) Radioimmunoassay for nicotine and cotinine. *IARC Sci Publ*, 293-9.
19. Godschalk, R.W., Maas, L.M., Van Zandwijk, N., van 't Veer, L.J., Breedijk, A., Borm, P.J., Verhaert, J., Kleinjans, J.C. and van Schooten, F.J. (1998) Differences in aromatic-DNA adduct levels between alveolar macrophages and subpopulations of white blood cells from smokers. *Carcinogenesis*, 19, 819-25.
20. Reddy, M.V. and Randerath, K. (1986) Nuclease P1-mediated enhancement of sensitivity of 32P-postlabeling test for structurally diverse DNA adducts. *Carcinogenesis*, 7, 1543-51.
21. Chang, T.K., Chen, J., Pillay, V., Ho, J.Y. and Bandiera, S.M. (2003) Real-time polymerase chain reaction analysis of CYP1B1 gene expression in human liver. *Toxicol Sci*, 71, 11-9.

22. Kim, J.H., Sherman, M.E., Curriero, F.C., Guengerich, F.P., Strickland, P.T. and Sutter, T.R. (2004) Expression of cytochromes P450 1A1 and 1B1 in human lung from smokers, non-smokers, and ex-smokers. *Toxicol Appl Pharmacol*, 199, 210-9.
23. Port, J.L., Yamaguchi, K., Du, B., De Lorenzo, M., Chang, M., Heerdt, P.M., Kopelovich, L., Marcus, C.B., Altorki, N.K., Subbaramaiah, K. and Dannenberg, A.J. (2004) Tobacco smoke induces CYP1B1 in the aerodigestive tract. *Carcinogenesis*, 25, 2275-81.
24. van Leeuwen, D.M., Gottschalk, R.W., van Herwijnen, M.H., Moonen, E.J., Kleinjans, J.C. and van Delft, J.H. (2005) Differential gene expression in human peripheral blood mononuclear cells induced by cigarette smoke and its constituents. *Toxicol Sci*, 86, 200-10.
25. Harju, T., Kaarteenaho-Wiik, R., Sirvio, R., Paakko, P., Crapo, J.D., Oury, T.D., Soini, Y. and Kinnula, V.L. (2004) Manganese superoxide dismutase is increased in the airways of smokers' lungs. *Eur Respir J*, 24, 765-71.
26. Shim, M.J., Kim, H.J., Yang, S.J., Lee, I.S., Choi, H.I. and Kim, T. (2002) Arsenic trioxide induces apoptosis in chronic myelogenous leukemia K562 cells: possible involvement of p38 MAP kinase. *J Biochem Mol Biol*, 35, 377-83.
27. Kim, Y.M., Reed, W., Lenz, A.G., Jaspers, I., Silbajoris, R., Nick, H.S. and Samet, J.M. (2005) Ultrafine carbon particles induce interleukin-8 gene transcription and p38 MAPK activation in normal human bronchial epithelial cells. *Am J Physiol Lung Cell Mol Physiol*, 288, L432-41.
28. Fang, J.Y. and Richardson, B.C. (2005) The MAPK signalling pathways and colorectal cancer. *Lancet Oncol*, 6, 322-7.
29. Liao, Y. and Hung, M.C. (2003) Regulation of the activity of p38 mitogen-activated protein kinase by Akt in cancer and adenoviral protein E1A-mediated sensitization to apoptosis. *Mol Cell Biol*, 23, 6836-48.
30. Ord, D. and Ord, T. (2005) Characterization of human NIPK (TRB3, SKIP3) gene activation in stressful conditions. *Biochem Biophys Res Commun*, 330, 210-8.
31. Osmak, M., Babic, D., Abramic, M., Milicic, D., Vrhovec, I., Skrk, J. (2001) Plasminogen activator inhibitor type 2: potential prognostic factor for endometrial carcinomas. *Neoplasma*, 48, 462-7.
32. Mengele, K., Harbeck, N., Reuning, U., Magdolen, V., and Schmitt, M. (2005) Tumor-associated prognostic factors of the plasminogen activator family: determination and clinical value of u-PA, t-PA, PAI-1, and PAI-2. *Hamostaseologie*, 25, 301-10.
33. Dickinson, J.L., Bates, E.J., Ferrante, A. and Antalis, T.M. (1995) Plasminogen activator inhibitor type 2 inhibits tumor necrosis factor alpha-induced apoptosis. Evidence for an alternate biological function. *J Biol Chem*, 270, 27894-904.
34. Senn, O., Russi, E.W., Imboden, M. and Probst-Hensch, N.M. (2005) α 1-Antitrypsin deficiency and lung disease: risk modification by occupational and environmental inhalants. *Eur Respir J*, 26, 909-17.
35. Kalsheker, N.A. (1994) Molecular pathology of alpha 1-antitrypsin deficiency and its significance to clinical medicine. *QJM*, 87, 653-58.
36. Yang, P., Bamlet, W.R., Sun, Z., Ebbert, J.O., Aubry, M.C., Krowka, M.J., Taylor, W.R., Marks, R.S., Deschamps, C., Swensen, S.J., Wieben, E.D., Cunningham, J.M., Melton, L.J. and de Andrade, M. (2005) Alpha1-antitrypsin and neutrophil elastase imbalance and lung cancer risk. *Chest*, 128, 445-52.
37. Nia, A.B., Maas, L.M., Brouwer, E.M., Kleinjans, J.C. and Van Schooten, F.J. (2000) Comparison between smoking-related DNA adduct analysis in induced sputum and peripheral blood lymphocytes. *Carcinogenesis*, 21, 1335-40.
38. Izzotti, A., Cartiglia, C., Longobardi, M., Balansky, R.M., D'Agostini, F., Lubet, R.A. and De Flora, S. (2004) Alterations of gene expression in skin and lung of mice exposed to light and cigarette smoke. *FASEB J*, 18, 1559-61.
39. Izzotti, A., Balansky, R.M., Cartiglia, C., Camoirano, A., Longobardi, M., De Flora, S. (2003) Genomic and transcriptional alterations in mouse fetus liver after transplacental exposure to cigarette smoke. *FASEB J*, 17, 1127-9.
40. Izzotti, A., Cartiglia, C., Longobardi, M., Bagnasco, M., Merello, A., You, M., Lubet, R.A., De Flora, S. (2004) Gene expression in the lung of p53 mutant mice exposed to cigarette smoke. *Cancer Res*, 64, 8566-72.
41. Bretz, F., Landgrebe, J. and Brunner, E. (2005) Multiplicity issues in microarray experiments. *Methods Inf Med*, 44, 431-7.

Chapter 4

Comparison of cigarette smoke-induced transcriptional responses in human airway epithelium and in blood cells

D.M. van Leeuwen¹
M.H.M. van Herwijnen¹
I. Jaspers²
J.C.S. Kleijnans¹
J.H.M. van Delft¹

¹ Department of Health Risk Analysis and Toxicology,
Maastricht University, Maastricht, the Netherlands

² UNC Center for Environmental Medicine, Asthma and Lung Biology,
Environmental Protection Agency Human Studies Facility,
Chapel Hill, North Carolina, USA

Abstract

The epithelial cell layer lining the bronchia and the lung suffers the highest exposure to respirable toxic and carcinogenic substances. Therefore toxic effects caused by respirable substances frequently manifest in the airway epithelium. Cigarette smoke is a complex mixture of respirable toxicants and carcinogens. The purpose of the present study therefore was to investigate cigarette smoke induced differential gene expressions in human airway epithelium cells, the primary target for cigarette smoke *in vivo*, and to compare the effects in these target cells with the responses that were previously described in surrogate target cells: *in vitro* in blood cells and in blood cells from smokers versus non-smokers. Using Agilent Human 22k oligonucleotide microarrays, modulated gene expressions were investigated in cultured normal human bronchial epithelial (NHBE) cells exposed to different concentrations of cigarette smoke condensate and for multiple exposure periods. Many genes were identified as significantly modulated in response to the condensate, in both a concentration as well as time-dependent manner. To generate a comprehensive overview of the transcriptomic changes, significant overrepresentation of biological processes and molecular functions was analyzed. Clear overrepresented processes and functions were found, although these differed according to concentration and exposure duration. Comparing the gene expression modulations with previous findings in blood cells revealed that the effects in NHBE agree to a certain extent with the main effects in PBMC *in vitro*. Therefore the genes and associated processes and functions adequately reflect expression changes in target as well as surrogate target cells. Specifically, *ATF4*, *SOD2*, *SERPINA1* and *CYP1B1* gene expression modulations by CSC were found to provide a link between responses of NHBE cells *in vitro* and blood cells *in vivo*. However, *CYP1B1* seems to be the only gene that was also significantly modulated in blood cells *in vitro* and *in vivo* and in NHBE.

Introduction

The epithelial cell layer of the human airways lining the bronchia and the lung suffers the highest exposure to toxic substances in ambient air and it is thus likely that toxic effects of respirable substances frequently primarily manifest in the airway epithelium. Bronchial epithelial cells of cigarette smokers are exposed to high concentrations of a complex mixture of toxicants and carcinogens, which are known to pose a major risk for the development of lung cancer. In order to deepen the knowledge about the response of target cells for lung cancer when exposed to toxic or carcinogenic insults, we investigated the transcriptional response of cultured human primary bronchial epithelial cells (NHBE), target cells for the development of lung cancer, to cigarette smoke exposure. An *ex vivo* study was performed as *in vivo* studies on lung epithelium among humans are very difficult to achieve because these would obviously require invasive procedures such as lung biopsies.

Cigarette smoke comprises an extensively studied complex mixture of over 4000 chemicals including more than 60 proven, probable and possible human carcinogens (according to IARC classifications) and is thus widely recognized as an important cause of cancer [1-3]. Cigarette smoking is a major risk factor for several diseases, especially for cancer, cardiovascular disease and lung conditions like chronic obstructive pulmonary disease (COPD) [4]. Currently, approximately 90% of all incident lung cancer cases are causally associated with cigarette smoking. Cigarette smoke is known to induce a wide range of genotoxic effects in target [5] as well as surrogate target tissue (blood cells), e.g. micronuclei, chromosome aberrations, gene mutations and DNA adducts, all of which are early indicators of cancer risk [6,7]. Furthermore, it has recently been observed that in human exposure to cigarette smoke affects gene expression in both target [5] tissue and surrogate target cells (blood) [8-13].

Spira *et al.* [10] are the first to have assessed the effects of smoking on the human lung epithelium transcriptome. Profiles of altered gene expression *in vivo* in human bronchial airway epithelial cells from smokers and non-smokers have been characterized using Affymetrix GeneChips. Cigarette smoke appeared to influence xenobiotic and redox-regulating genes as well as genes involved in inflammation. Fields *et al.* [12] have investigated expression changes by means of quantitative real-time PCR for four specific genes, involved in DNA damage response, oxidative stress, inflammation and cell cycle regulation, in NHBE cells following *in vitro* exposure to cigarette smoke condensates. They found that CSC significantly induced the expression of each gene significantly. However, there have been no reports up till now on the genome-wide exploration of gene expression alterations after exposure to cigarette smoke (condensate) *in vitro*, to grasp a view on the early, overall effects of this carcinogenic exposure on the transcriptome level in an experimental setting.

In addition, cigarette smoke-induced differential gene expression has been studied in blood cells [8,13,14]. An *in vitro* study on differential gene expression in peripheral blood mononuclear cells (PBMC) exposed to cigarette smoke condensate was performed by us, as well as an *in vivo* investigation of altered gene expression in blood from monozygotic cigarette smoking-discordant twins (**Chapters 2 and 3**) [8,14]. In both studies sets of genes were identified that might be candidate biomarkers for cigarette

smoke exposure and provided mechanistic information on the effects occurring in surrogate target cells. Also, a study by Lampe *et al.* investigated *in vivo* gene expression in smokers versus non-smokers [13]. Thirty-six genes were reported that significantly correlated with plasma cotinine levels.

A common issue in studies that monitor (the biological effects of) human exposure to chemicals, is the interpretation of data from surrogate target tissue with respect to their biological relevance. Due to practical and ethical limitations, human biomonitoring studies can not be conducted using target tissue or cells like primary airway (epithelial) cells. Therefore, blood cells are the surrogate target tissue of choice in biomonitoring studies.

The purpose of the present study was to investigate cigarette smoke induced differential gene expressions in human airway epithelium cells (the primary targets *in vivo*) and to compare the effects in these target cells with the responses that previously have been described *in vitro* and smokers *in vivo* in blood cells (surrogate target cells) [8,14]. To achieve this, NHBE cells were exposed to a condensate of mainstream cigarette smoke, at three concentrations and during three exposure periods. Using Agilent Human oligonucleotide microarrays, containing more than 22.000 probes for genes, modulation of gene expression was investigated. These data were also used to identify possible affected biological processes and molecular functions.

Materials and methods

Cell culture and exposures

Primary normal human bronchial epithelial (NHBE) cells were obtained from healthy nonsmoking adult volunteers by cytologic brushing at bronchoscopy and prepared for culture as described previously [15], according to an acquisition protocol for lung cells as reviewed by the University of North Carolina Institutional Review Board. The NHBE cells were expanded to passage 2 in bronchial epithelial growth medium (BEGM, Cambrex Bioscience Walkersville, Inc., Walkersville, MD) and then plated on collagen-coated filter supports with a 0.4 μm pore size (Trans-CLR; Costar, Cambridge, MA) and cultured in a 1:1 mixture of bronchial epithelial cell basic medium (BEBM) and DMEM-H with Single-Quot supplements (Cambrex), bovine pituitary extracts (13 mg/ml), bovine serum albumin (BSA, 1.5 $\mu\text{g/ml}$), and nystatin (20 units). All-trans retinoic acid was added to the medium upon confluence. Air liquid interface (ALI) culture conditions, *i.e.* removal of the apical medium, were created to promote differentiation. Mucociliary differentiation was achieved after 18-21 days post-ALI.

In vitro exposures

NHBE cells were exposed to several concentrations of cigarette smoke condensate (CSC; stock contains 100 mg total particulate matter CSC/ml DMSO [8]) diluted in culture medium (final concentration of DMSO is 0.3% v/v). Volumes of 100 μl were applied to the apical side of the cell cultures, mimicking the *in vivo* exposure as close as possible. Concentrations of CSC ranged from 0.1 to 300 $\mu\text{g/ml}$ for the real-time PCR analysis and from 3 to 30 $\mu\text{g/ml}$ for the microarray analysis. After treatments, cells were harvested

and homogenized in Trizol[®] (Life Technologies, Carlsbad, CA, USA) for RNA isolation. Total RNA was isolated according to the manufacturer's instructions and stored in nuclease-free H₂O at -80°C until cDNA synthesis. Six independent treatments of NHBE cells from six different donors were conducted: two for real-time PCR analysis and four for the microarray studies.

Real-time PCR analysis of gene expression

In order to follow interleukin 6 (*IL6*) and 8 (*IL8*) gene expressions after CSC exposure, quantitative real time-PCR was performed as described previously [16]. These genes were chosen as they are known to reflect an inflammatory (defense) response and are regulated by similar exposures [8,17]. The concentration-dependent responses of these genes served as an indicator for the CSC concentrations to be chosen for the subsequent microarray-based gene expression experiments. Glyceraldehyde-3-phosphate dehydrogenase (*GAPD*) was included as internal control gene.

IL-6:

probe, 5'-FAM-TGTTACTCTTGTTACATGTCTCCTTTCTCAGGGCT-TAMRA-3';

sense, 5'-GGTACATCCTCGACGGCATCT-3';

antisense, 5'-GTGCCCTTTGCTGCTTTTACAC-3';

IL-8:

probe, 5'-FAM-CCTTGGCAAACTGCACCTTCACACA-TAMRA-3';

sense, 5'-TTGGCAGCCTTCCTGATTTTC-3';

antisense, 5'-TATGCACTGACATCTAAGTTCTTTAGCA-3'

GAPD:

probe, 5'-JOE-CAAGCTTCCCGTTCTCAGCC-TAMRA-3';

sense, 5'-GAAGGTGAAGGTCGGAGTC-3';

antisense, 5'-GAAGATGGTGATGGGATTTTC-3'.

Microarray analysis of multi-gene expression

Total RNA was further purified using the RNeasy system (Qiagen, Hilden, Germany). For hybridization onto Agilent Human 22k oligonucleotide arrays (G4110B; Agilent Technologies, Palo Alto, CA), one µg of total RNA was used in the Agilent Low Input RNA Linear Amplification method, generating Cyanine5-CTP (Perkin Elmer, Boston, MA, USA) or Cyanine3-CTP labeled cRNA, according to the manufacturer's instructions. Cyanine incorporation efficiency into the cRNA was determined by UV-absorbance and cRNAs showing a cyanine incorporation of 7-20 pmol/µg cRNA were selected for further processing.

Figure 1. Hybridization design of the study, per donor and per time point. CSC treated samples and the vehicle control treated sample were labeled with Cy5 and competitively hybridized against a second aliquot of the vehicle control treated sample, labeled with Cy3. Each arrow indicates an array.

Hybridizations were conducted according to the design shown in Figure 1. Equal amounts of test and reference sample (always donor- and time-matched vehicle controls) were competitively hybridized onto Agilent 22k Human oligonucleotide microarrays in a hybridization oven at 60°C for 17 hours. Slides were washed according to the manufacturer's instructions with SSC-based wash buffers and finally dipped in Drying and Stabilization buffer (Agilent) diluted with acetonitrile (Drying and Stabilization buffer: acetonitrile = 2:1). The arrays were scanned within one hour after washing on a ScanArrayExpress scanner (Perkin Elmer, Boston, MA, USA), which generated separate Cy5 and Cy3 tif-images (10 µm pixel size). Laser power was set to 100% and the photomultiplier gain to 60% for Cy3 and 55% for Cy5 scans. The resulting image files were uploaded pair wise into ImaGene (BioDiscovery, Marina del Rey, CA, USA) for translation into raw data files. Further data preparations and analyses were performed in GeneSight (BioDiscovery). Data transformations included in the following order: local background subtraction, omission of bad spots (defined automatically by ImaGene based on abnormal spot shapes), elimination of spots with low signal intensities (<20 arbitrary units in both channels), log (base=2) transformation, locally weighted scatter plot smoothing normalization (LOWESS; global, 0.2 smoothing factor) and calculation of the difference of the treated test sample with the reference sample.

Data analysis and data mining

All analyses were conducted in GeneSight using average data for the 4 donors (confidence analyses) or using the individual data (t-tests). Confidence analyses were performed at >99% confidence level and >0.1 regulation level for each CSC concentration at each time point, in order to filter out the genes that most likely responded to the treatments. These filtered gene lists were then analyzed by 2-tailed Student t-tests of the treated-control hybridizations versus the control-control hybridizations, in order to identify statistically-significantly responding genes per concentration within a time point (at $p < 0.001$, $p < 0.01$ and $p < 0.05$). The lists of significantly responding genes were used for further data mining of possibly affected biological processes or molecular functions. To

this end, gene lists of interest were analyzed using the online tool Expression Analysis Systematic Explorer (EASE, <http://apps1.niaid.nih.gov/David>) [18] to explore significant enrichment or over-representation of biological processes or molecular functions (EASE score <0.05).

Results

Real-time PCR analysis of IL6 and IL8 gene expression

To select the concentration range for the NHBE cell treatments for gene expression profiling by microarray analysis, initially a concentration-response experiment was performed. NHBE cells originating from two different donors were exposed to a broad concentration series of CSC (0.1 to 300 µg/ml) for 18 hours or to the vehicle control DMSO (0.3% v/v). Cell viability was determined by microscopic inspection of the cells together with measurement of cilia beat frequencies (data not reported). Both were not observed to be impaired by the exposures. Expression responses of the *IL6* and *IL8* genes were analyzed by real-time PCR. Final concentrations of 3, 10 and 30 µg/ml were selected for the subsequent microarray gene expression profiling experiments based on the response of these genes (Figure 2) to CSC.

Figure 2. *IL6 and IL8 gene expression in two donors (A and B) measured by real-time PCR in NHBE cells after 18 hours of exposure to a concentration series of CSC. Values are normalized to GAPD and relative to the vehicle control (DMSO).*

Microarray analysis of multi-gene expression

Transcriptomic responses after exposure periods of 2, 24 or 96 hours to 0, 3, 10 or 30 $\mu\text{g/ml}$ CSC were measured by microarray technology. Independent studies were conducted on cells from 4 different donors, in order to cope with inter-experimental and inter-individual variations. The numbers of significantly modulated genes are depicted in Figure 3, and those affected by all three concentrations are reported in the intersections of the three rings per diagram. For every gene in the intersections, the direction of effect was similar for each of the three concentrations.

With regard to time-related responses to the CSC exposures, the incubation period of 2h was observed to cause the most genes to respond whereas after 24 and 96 rather comparable numbers of genes were modulated (Figure 3). The highest numbers of modulated gene expressions were observed at the highest CSC concentration at all time points, whereas lower concentrations altered the expression of grossly similar numbers of genes. Figure 4 shows the expression values for the 13 genes from the intersections of Figure 3, which are regarded as the main CSC-induced responses in the NHBE cells. The largest effects, in terms of magnitude of the expression difference ($\log_2$), were observed after the shortest exposures. The gene most frequently affected in all experiments was *CYP1A1*. The largest effects were observed for *CYP1A1* and *CYP1B1*, which were in the range of 8- to 16-fold up-regulation (3 to 4 on a log base2 scale, Figure 4). Overall, the vast majority of genes appeared up-regulated.

For comparison of the findings in the NHBE cells with previous observations in blood cells, the analyses were restricted to genes that were present on both microarray platforms used; Agilent Human 22k arrays in the current study and PHASE-1 Human Tox 600 (Santa Fe, USA) in the previous blood cell studies.

Comparison of transcriptional response in airway and blood cells

Figure 3. Numbers of genes significantly responding in NHBE cells to different concentrations of CSC after 2, 24 and 96 h of exposure at different p-values ($p<0.001$, $p<0.01$ and $p<0.05$ from top to bottom). Values in the centers of each diagram (intersections of the three rings) represent the genes responding significantly at all concentrations.

Figure 4. Differential expressions of the genes from the intersections of Figure 3 over time, relative to the vehicle control (DMSO). Values are expressed as mean $\pm$ standard error of the mean for 4 different donors.

Figure 6. Gene expression changes observed in blood cells from smokers versus non-smokers and in NHBE cells exposed to CSC for 96 h. Since in both studies different array platforms were used, only genes present on both platforms are included. An asterisk indicates statistical significance ($p < 0.05$). *CYP1B1* gene expression in the smokers' blood cells was measured by means of real-time PCR.

Comparison of NHBE with PBMC *in vitro*

For comparison of CSC-induced transcriptomic responses in target cells with surrogate target tissue, the current observations in NHBE cells were compared to the main modulated CSC-responsive genes from a previously conducted *in vitro* study on PBMC (**Chapter 2**) (Figure 5); the first 13 genes from the left of the figure. Since for both studies different array platforms were used, only genes present on both platforms are included in the comparison. The (altered) gene expressions of *IL1B*, *IL6*, *CXCL2*, *ENO1*, *VEGF* and *CYP1B1* appeared to be similar in both cell types with respect to the direction of regulation. However, the overall magnitudes of the effects in the NHBE cells were smaller than the effects in PBMC. Only the expression of *CXCL2* appeared significantly affected in NHBE cells after exposure to 10 $\mu\text{g/ml}$ ($p < 0.05$).

Comparison of NHBE with blood cells from smokers *in vivo*

In addition, the responses in NHBE cells were compared to the findings of a previous *in vivo* study on the transcriptomic effects of cigarette smoke exposure in blood cells from smokers versus non-smokers. Also here, only genes present on both array platforms are included in the comparison. NHBE gene expressions that were included in the comparison are those from the 96 h exposure experiments as this rather chronic *in vitro* exposure (four daily dosages of CSC were applied) presumably best resembles the exposure of smokers. Gene expressions that have been found to significantly differ between smokers and non-smokers were investigated in NHBE cells (Figure 6). The closest resemblance was observed for the *CYP1B1* gene, with regard to both the magnitude and the significance of the induced effect. Although not significantly altered in NHBE cells, *ATF4*, *PXN*, *SERPINA1* and *SOD2* gene expressions also showed to some extent similar modulations in smokers' blood and the target cells.

Analysis of overrepresentation of functionally related gene clusters

Using the software tool EASE, overrepresentation was investigated of responding genes in biological processes and molecular functions defined by Gene Ontologies. The lists of possibly affected biological processes and molecular functions are presented in Figure 7 (EASE score <0.05). Per concentration and exposure duration, the five most affected processes or functions are shown. The highest exposure over the longest time (96 h) seems to affect the most processes and functions. Besides the five that are presented in Figure 7, this exposure affected many more processes and functions, including oxidoreductase activity, histogenesis, organogenesis and morphogenesis, immune and inflammatory response and musculoskeletal movement. Overall observations are the enrichment of molecular functions such as oxidoreductase activity and biological processes associated with inflammatory response. Growth factor activity specifically appeared affected after CSC exposure for 2 hours.

Figure 7. Biological processes and molecular functions significantly affected in NHBE cells after exposure to different concentrations of CSC as reported by EASE. Values in brackets represent EASE scores.

Discussion

The present study provides a genome-wide analysis of gene expression changes in cultured NHBE cells *in vitro* with the purpose to compare CSC-induced differential gene expression in target tissue with that previously described in surrogate target cells, namely blood cells. NHBE cells originating from four independent, healthy and non-smoking donors were cultured in a transwell membrane system which resembles the *in vivo* airway epithelium as it represents basal, mucous secreting, ciliated and non-ciliated cells. NHBE cells were exposed to three concentrations of cigarette smoke condensate for 2, 24 or 96 hours. It was *a priori* decided to chose the highest concentration of CSC to be applied in the current study, based on a 'nick' in the gene expression response at the highest concentration as observed in the previous PBMC study (**Chapter 2**), where gene expression was frequently observed to change from up- to down-regulation or to decrease in expression [8]. With this knowledge, the initial concentration-response experiment was performed. Figure 2 shows the expression response with increasing concentrations of CSC for both donors A and B separately as donor A was observed to display a peak increase of both *IL6* and *IL8* expression at 10 µg/ml and donor B at 100 µg/ml. Therefore, 30 µg/ml was chosen to be the highest concentration to be applied for the microarray gene expression analysis. Differential gene expression of exposed versus non-exposed cells was analyzed using Agilent oligonucleotide microarrays. In order to reduce of the number of false-positive calls when performing statistical tests for multiple genes, first genes that were most likely to show a true response were selected by means of confidence analysis; these lists of genes then served as the basis for further Students t-tests to retrieve statistically-significant CSC-responsive genes.

First it was examined to what extent genes in NHBE cells responded consistently and significantly to the different CSC exposures after the different exposure periods, with respect to time- and concentration-responsiveness. Although an upward time-response relationship might be expected, as a longer exposure can cause a heavier burden with more genes to respond, this was not observed. The largest effect of CSC exposure, in terms of numbers of genes significantly modulated, was observed at the 2 h treatment (Figure 3). The studied primary normal bronchial epithelial cells encompass a specific type of cells, which are first and foremost exposed to respirable chemicals. As bronchial epithelial cells are constantly exposed to air, it would be plausible explanation for these cells to act to xenobiotic stimuli rapidly by modulation of gene expression; a possible intrinsic characteristic. To our knowledge, at present there are no publications on similar studies that report the same trend in terms of numbers of genes responding time-dependently to chemical carcinogen exposure in primary cells, as analyzed by microarray technology.

Concentration-effect relationships were observed irregardless of the exposure duration, as the highest CSC concentration consequently caused the largest numbers of genes to significantly respond (Figure 3). Thirteen genes that showed significant modulation at each of the three concentrations (as shown in the intersections of the three diagrams in Figure 3) were studied in more detail. These genes are regarded as the overall most relevant genes that mark the response of NHBE cells to CSC *in vitro*. Of these, the genes encoding cytochrome P450 family members 1A1 (*CYP1A1*) and 1B1 (*CYP1B1*)

showed the largest effects (Figure 4). Except for *CYP1B1* after 96 h of exposure to 3 µg/ml CSC, *CYP1A1* and *CYP1B1* also responded significantly to all exposures, *i.e.* to all concentrations and each exposure duration. The majority of the 13 genes showed a concentration-response relationship (Figure 4). Some of these genes can be related to exposure to cigarette smoke or its effects. *CYP1A1* and *CYP1B1* are known to be involved in the bio-activation of pro-carcinogens such as polycyclic aromatic hydrocarbons (PAHs) present in cigarette smoke. Detection of (induction of) expression of these genes is consistent with literature [19,20]. Affected regulation of *CYP1B1* and *CYP1A1* by xenobiotic exposure has been reported in liver and lung from human smokers *in vivo* [10,21,22] and in lung cells exposed to cigarette smoke *in vitro* [23]. Moreover, *CYP1B1* gene expression has been observed to be elevated in peripheral blood cells from lung cancer patients [24]. Interleukins 1A and 1B (*IL1A* and *IL1B*) are cytokines and mediate inflammatory or immune responses. *IL1B* is also involved in a variety of cellular activities, including cell proliferation, differentiation, and apoptosis. Amphiregulin (*AREG*) encodes a protein that belongs to the epidermal growth factor family. Binding of *AREG* to the epidermal growth factor receptor in response to tobacco smoke exposure has been found to stimulate proliferation of normal lung epithelial cells *in vitro* [25].

For a comprehensive, generic overview of the response of NHBE cells to CSC *in vitro*, analyses of significant enrichment (or overrepresentation) of modulated genes in specific biological processes or molecular functions were performed using the online Expression Analysis Systematic Explorer (EASE). Figure 7 shows the five most significantly affected processes and functions for each concentration per exposure duration and their EASE scores. These analyses revealed different biological processes or molecular functions affected after 2, 24 or 96 h. Processes and functions affected after the 2 h exposures were mainly related to growth factor activity, (interleukin-1) receptor binding and cytokine activity, which indicate a rather acute defensive response. Amine/polyamine transporter activity, antigen presentation/processing and transferase activity were the most affected processes after 24 h exposures to 3, 10 and 30 µg/ml CSC respectively. After the more chronic exposure of 96 h, exposures to the respective concentrations CSC returned G-protein coupled receptor activity, organismal physiologic process and epidermal differentiation as the most overrepresented processes/functions (Figure 7). The most significantly overrepresented processes/functions seem specific to each concentration and each exposure duration. At this moment, we cannot provide an explanation why effects differ so much between the different dose levels, especially at the 24 and 96 h time points.

The observations in NHBE cells were also compared with previous findings in peripheral blood cells, the latter from both *in vitro* as well from *in vivo* studies as described in **Chapters 2 and 3** [8,14], thus comparing the effects of CSC on target and surrogate target cells. Since different array platforms were used in the NHBE and PBMC studies, the comparisons are limited to those genes that are present on both platforms. Some similarities were found in genomic responses to CSC exposure in NHBE and blood cells (Figures 5 and 6).

Starting with the *in vitro* NHBE - PBMC comparison, expressions of the genes that were found to most sensitively respond to CSC in PBMC were investigated in the NHBE cells, as the modulations of these genes were the specific outcomes of the PBMC study. The expression differences of exposed versus non-exposed PBMC and NHBE cells for these genes are presented in Figure 5. *CYP1A1* and *CYP1B1* expressions are

depicted as these were the most affected genes in the NHBE cells at this time point. Expression modulations that were found to be significant in both PBMC and NHBE cells were *CXCL2* (in NHBE at the medium concentration of 10 µg/ml CSC) and *CYP1B1* (at all concentrations CSC except for 3 µg/ml in NHBE cells). Although in some other cases similarity was observed between PBMC and NHBE cells with regard to the direction of expression, the overall magnitude of the effects in the airway cells was smaller than in PBMC. The genes that showed (most) resemblance in terms of direction of regulation and concentration-responsiveness were *IL1B*, *IL6*, *ENO1*, *VEGF* and *CYP1B1*. Biological processes and molecular functions as revealed by EASE analyses in both PBMC and NHBE are inflammation-related processes and functions like cytokine activity and interleukin-1 receptor binding [8]. However, for the NHBE cells, but not for PBMC, EASE reported additional functions as affected (Figure 7). In conclusion, the effects found in NHBE confirm to a certain extent the main effects found in PBMC induced by exposure to CSC *in vitro* and therefore these genes and associated processes and functions adequately reflect expression changes in target as well as surrogate target cells.

Furthermore, *in vitro* target tissue (NHBE cells) responses were compared to effects due to smoking described in surrogate target cells (blood cells), from smoking-discordant monozygotic twin pairs [14]. For this comparison, the differential expression of genes in NHBE cells after 96 h of exposure was used because of the more chronic character of this exposure, which probably best reflects the chronic exposures in smokers *in vivo* (Figure 6). For the genes found deregulated in smokers versus non-smokers, no statistically significant expression changes were found in NHBE cells. However, activating transcription factor 4 (*ATF4*), superoxide dismutase 2 (*SOD2*), alpha-1-antitrypsin (*SERPINA1*) and *CYP1B1* expressions were regulated in the same direction in both cell types, *i.e.* up-regulation. Thus *ATF4*, *SOD2*, *SERPINA1* and *CYP1B1* gene expression modulation by CSC provides a link between response of target cells *in vitro* and surrogate target cells *in vivo*. *SERPINA1* is a protease inhibitor, deficiency of which is associated with emphysema [26]. Recently, it has been associated with lung cancer risk [27]. The biological relevance of differential *ATF4*, *SOD2* and *CYP1B1* gene expressions to (the effects of) cigarette smoking have been described above and previously [14].

Relating our findings to other studies, some resemblances are appointed, although few reports are available on effects of tobacco smoke exposure on primary human cells at the genome level. In a study on the effects of cigarette smoke on the *in vivo* airway transcriptome, tobacco smoke-induced genes have been identified which were mainly involved in biological processes like induction of xenobiotic metabolism, regulation of oxidative stress and glutathione metabolism, secretion and repression of inflammation [10]. Our data in part adhere to these observations as oxidoreductase activity was seen to be influenced. However, we did not observe a decrease in the expression of inflammatory genes. This might be due to the short term exposures that were investigated *in vitro* in the present study as opposed to the long term *in vivo* exposure in smokers. Hellerman *et al.* investigated the inflammatory response of NHBE cells to a cigarette smoke condensate *in vitro* and found elevation of *IL1B*, *IL6* and *IL8* gene expressions, which is confirmed by our study [11]. The magnitude of the effects was somewhat higher in that study however, which might be due to a shorter exposure time (30 minutes) and/or to their use of CSC prepared from Kentucky 1R3F reference cigarettes instead of 2R4F which was used in our studies; 1R3F cigarettes have a higher tar and nicotine yield than 2R4F, whereas 2R4F cigarettes represent the average full-flavor 'low tar'

cigarette [12]. Recently, Fields *et al.* reported that CSC, from the same Kentucky reference cigarettes as used in the present study, alters the expression of a selected set of five genes in NHBE *in vitro*, involved in DNA damage, cell cycle regulation, oxidative stress and inflammation (*GADD45*, *p21 (WAF/CIP1)*, *COX2* and *IL8* respectively) [12]. Their CSC concentrations were comparable to those in our study, but only a 3 hour exposure period was investigated. Our results (at the 2h exposure time point) confirm the induction of all five genes by CSC, although only *GADD45* showed statistical significance (data not shown). Furthermore, other studies have reported transcriptomic responses of lung cells following exposure to environmental agents and carcinogens other than cigarette smoke. These studies were focused on A549 or HPL1A cells however, which are not primary human airway cells but malignant or immortalized lung epithelial cells [28], investigated the effects on gene expression by other compounds such as TCDD, ultrafine particles [17] and diesel exhaust [15] or were in diseased study subjects [29,30].

In conclusion, we identified many genes to be differentially expressed in cultured NHBE cells (target cells for lung cancer) following exposure to CSC. Identification of biological processes and molecular functions that appear affected in the human normal airway epithelium after *in vitro* treatment with CSC provides information on the earliest events occurring in the lung after exposure to such a complex carcinogenic mixture. We compared NHBE responses with the expression changes observed in our previous studies on PBMC (the surrogate target cells) following exposure to cigarette smoke *in vitro* as well as *in vivo*. We can appoint several genes of which the expression alters in response to CSC exposure in target as well as surrogate target cells, however *CYP1B1* seems to be the only gene that was significantly modulated in all the three studies that were compared.

Acknowledgements

The authors thank the donors for donating the NHBE cells, L. Brighton for cell culturing and Dr. J. Carson for assistance on cilia beat frequency measurements.

References

- Smith, C.J., Perfetti, T.A., Garg, R. and Hansch, C. (2003). IARC carcinogens reported in cigarette mainstream smoke and their calculated log P values. *Food Chem. Toxicol* 41, 807-817.
- Hecht, S.S. (2002). Cigarette smoking and lung cancer: chemical mechanisms and approaches to prevention. *Lancet Oncol* 3, 461-469.
- Rodgman, A., Smith, C.J. and Perfetti, T.A. (2000). The composition of cigarette smoke: a retrospective, with emphasis on polycyclic components. *Hum Exp Toxicol* 19, 573-595.
- IARC IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans. Volume 38. Tobacco smoking, IARC, Switzerland, 1986.
- Liou, S.H., Lung, J.C., Chen, Y.H., Yang, T., Hsieh, L.L., Chen, C.J. and Wu, T.N. (1999). Increased chromosome-type chromosome aberration frequencies as biomarkers of cancer risk in a blackfoot endemic area. *Cancer Res* 59, 1481-1484.
- De Flora, S., D'Agostini, F., Balansky, R., Camoirano, A., Bennicelli, C., Bagnasco, M., Cartiglia, C., Tampa, E., Longobardi, M.G., Lubet, R.A. and Izzotti, A. (2003). Modulation of cigarette smoke-related end-points in mutagenesis and carcinogenesis. *Mutat. Res* 523-524, 237-252.
- DeMarini, D.M. (2004). Genotoxicity of tobacco smoke and tobacco smoke condensate: a review. *Mutat Res* 567, 447-474.
- van Leeuwen, D.M., Gottschalk, R.W., van Herwijnen, M.H., Moonen, E.J., Kleinjans, J.C. and van Delft, J.H. (2005). Differential gene expression in human peripheral blood mononuclear cells induced by cigarette smoke and its constituents. *Toxicol Sci* 86(1), 200-210.
- Spira, A., Schembri, F., Beane, J., Shah, V., Liu, G. and Brody, J.S. (2004). Impact of cigarette smoke on the normal airway transcriptome. *Chest* 125, 115S.
- Spira, A., Beane, J., Shah, V., Liu, G., Schembri, F., Yang, X., Palma, J. and Brody, J.S. (2004). Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci U S A* 101, 10143-10148.
- Hellermann, G.R., Nagy, S.B., Kong, X., Lockey, R.F. and Mohapatra, S.S. (2002). Mechanism of cigarette smoke condensate-induced acute inflammatory response in human bronchial epithelial cells. *Respir. Res* 3, 22.
- Fields, W.R., Leonard, R.M., Odom, P.S., Nordskog, B.K., Ogden, M.W. and Doolittle, D.J. (2005). Gene expression in normal human bronchial epithelial (NHBE) cells following in vitro exposure to cigarette smoke condensate. *Toxicol Sci* 86, 84-91.
- Lampe, J.W., Stepaniants, S.B., Mao, M., Radich, J.P., Dai, H., Linsley, P.S., Friend, S.H. and Potter, J.D. (2004). Signatures of environmental exposures using peripheral leukocyte gene expression: tobacco smoke. *Cancer Epidemiol. Biomarkers Prev* 13, 445-453.
- van Leeuwen, D.M., van Agen, B., Gottschalk, R.W.H., Vlietinck, R., Gielen, M., van Herwijnen, M.H.M., Maas, L., Kleinjans, J.C.S. and van Delft, J.H.M. (2006). Cigarette smoke-induced differential gene expression in blood cells from monozygotic twin pairs. *Carcinogenesis* (in press).
- Jaspers, I., Cieniewicz, J.M., Zhang, W., Brighton, L.E., Carson, J.L., Beck, M.A. and Madden, M.C. (2005). Diesel exhaust enhances influenza virus infections in respiratory epithelial cells. *Toxicol Sci* 85, 990-1002.
- Jaspers, I., Zhang, W., Fraser, A., Samet, J.M. and Reed, W. (2001). Hydrogen peroxide has opposing effects on IKK activity and IkappaBalpha breakdown in airway epithelial cells. *Am J Respir Cell Mol Biol* 24, 769-777.
- Kim, Y.M., Reed, W., Lenz, A.G., Jaspers, I., Silbajoris, R., Nick, H.S. and Samet, J.M. (2005). Ultrafine carbon particles induce interleukin-8 gene transcription and p38 MAPK activation in normal human bronchial epithelial cells. *Am J Physiol Lung Cell Mol Physiol* 288, L432-441.
- Hosack, D.A., Dennis, G., Jr., Sherman, B.T., Lane, H.C. and Lempicki, R.A. (2003). Identifying biological themes within lists of genes with EASE. *Genome Biol* 4, R70.
- Wei, C., Caccavale, R.J., Kehoe, J.J., Thomas, P.E. and Iba, M.M. (2001). CYP1A2 is expressed along with CYP1A1 in the human lung. *Cancer Lett* 171, 113-120.
- Spivack, S.D., Hurteau, G.J., Reilly, A.A., Aldous, K.M., Ding, X. and Kaminsky, L.S. (2001). CYP1B1 expression in human lung. *Drug Metab Dispos* 29, 916-922.
- Chang, T.K., Chen, J., Pillay, V., Ho, J.Y. and Bandiera, S.M. (2003). Real-time polymerase chain reaction analysis of CYP1B1 gene expression in human liver. *Toxicol Sci* 71, 11-19.
- Kim, J.H., Sherman, M.E., Curriero, F.C., Guengerich, F.P., Strickland, P.T. and Sutter, T.R. (2004). Expression of cytochromes P450 1A1 and 1B1 in human lung from smokers, non-smokers, and ex-smokers. *Toxicol Appl Pharmacol* 199, 210-219.

Comparison of transcriptional response in airway and blood cells

23. Han, W., Pentecost, B.T., Pietropaolo, R.L., Fasco, M.J. and Spivack, S.D. (2005). Estrogen receptor alpha increases basal and cigarette smoke extract-induced expression of CYP1A1 and CYP1B1, but not GSTP1, in normal human bronchial epithelial cells. *Mol Carcinog* 44, 202-211.
24. Wu, M.F., Wu, W.J., Chang, G.C., Chen, C.Y., Hu, S.W., Tsai, W.T., Lee, H. and Lin, P. (2004). Increased expression of cytochrome P4501B1 in peripheral leukocytes from lung cancer patients. *Toxicol Lett* 150, 211-219.
25. Lemjabbar, H., Li, D., Gallup, M., Sidhu, S., Drori, E. and Basbaum, C. (2003). Tobacco smoke-induced lung cell proliferation mediated by tumor necrosis factor alpha-converting enzyme and amphiregulin. *J Biol Chem* 278, 26202-26207.
26. Kalsheker, N.A. (1994). Molecular pathology of alpha 1-antitrypsin deficiency and its significance to clinical medicine. *Qjm* 87, 653-658.
27. Yang, P., Bamlet, W.R., Sun, Z., Ebbert, J.O., Aubry, M.C., Taylor, W.R., Marks, R.S., Deschamps, C., Swensen, S.J., Wieben, E.D., Cunningham, J.M., Melton, L.J. and de Andrade, M. (2005). Alpha1-antitrypsin and neutrophil elastase imbalance and lung cancer risk. *Chest* 128, 445-452.
28. Martinez, J.M., Afshari, C.A., Bushel, P.R., Masuda, A., Takahashi, T. and Walker, N.J. (2002). Differential toxicogenomic responses to 2,3,7,8-tetrachlorodibenzo-p-dioxin in malignant and nonmalignant human airway epithelial cells. *Toxicol Sci* 69, 409-423.
29. Ning, W., Li, C.J., Kaminski, N., Feghali-Bostwick, C.A., Alber, S.M., Di, Y.P., Otterbein, S.L., Song, R., Hayashi, S., Zhou, Z., Pinsky, D.J., Watkins, S.C., Pilewski, J.M., Sciruba, F.C., Peters, D.G., Hogg, J.C. and Choi, A.M. (2004). Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease. *Proc Natl Acad Sci U S A* 101, 14895-14900.
30. Spira, A., Beane, J., Pinto-Plata, V., Kadar, A., Liu, G., Shah, V., Celli, B. and Brody, J.S. (2004). Gene expression profiling of human lung tissue from smokers with severe emphysema. *Am J Respir Cell Mol Biol* 31, 601-610.

Chapter 5

Genome-wide differential gene expression in children exposed to air pollution in the Czech Republic

D.M. van Leeuwen¹
M.H.M. van Herwijnen¹
M. Pedersen²
L.E. Knudsen²
M. Kirsch-Volders³
R.J. Sram⁴
Y.C.M. Staal¹
E. Bajak⁵
J.H.M. van Delft¹
J.C.S. Kleinjans¹

¹ Department of Health Risk Analysis and Toxicology, Maastricht University, Maastricht, Netherlands

² Department of Environmental and Occupational Health, University of Copenhagen, Denmark

³ The Faculty of Science, Laboratory of Cell Genetics, Vrije Universiteit Brussel, Brussel, Belgium

⁴ Institute of Experimental Med. AS CR & Health Institute Central Bohemia, Prague, Czech Republic

⁵ Karolinska Institutet, Department of Environmental Medicine, Stockholm, Sweden

Abstract

The Teplice area in the Czech Republic is a mining district where elevated levels of air pollution including airborne carcinogens, have been demonstrated, especially during winter time. This environmental exposure can impact human health; in particular children may be more vulnerable. To study the impact of air pollution in children at the transcriptional level, peripheral blood cells were subjected to whole genome response analysis, in order to identify significantly modulated biological pathways and processes as a result of exposure. Using genome-wide oligonucleotide microarrays, we investigated differential gene expression in children from the Teplice area ($n=23$) and compared them with children from the rural control area of Prachatice ($n=24$). In an additional approach, individual gene expressions were correlated with individual peripheral blood lymphocyte micronuclei frequencies, in order to evaluate the linkage of individual gene expressions with an established biomarker of effect that is representative for increased genotoxic risk. Children from the Teplice area showed a significantly higher average micronuclei frequency than Prachatice children ($p=0.023$). For considerable numbers of genes, the expression differed significantly between the children from the two areas. Amongst these genes, considerable numbers of genes were observed to correlate significantly with the frequencies of micronuclei. The main biological process that appeared significantly affected overall was nucleosome assembly. This suggests an effect of air pollution on the primary structural unit of the condensed DNA. In addition, several other pathways were modulated. Based on the results of this study, we suggest that transcriptomic analysis represents a promising biomarker for environmental carcinogenesis.

Introduction

Exposure to xenobiotics has quite different impact on children than it has on adult humans. It has been hypothesized that infants are more susceptible to exposures to environmental toxicants and carcinogens [1-3]. Per kilogram of body weight, children have higher daily intake of air and food than adults, which is likely to result in a higher endogenous exposure to (air) pollutants [3]. Furthermore, in general, children spend more time outdoors showing higher physical activity, which also increases exposure. In addition, children are considered more susceptible to toxicants than adults since they exhibit physiological differences related to growth, development and immaturity of organs, resulting in different features of target organs. Children also exhibit immaturity of metabolic pathways; as a consequence the rates of distribution of chemicals and their ability to detoxify and excrete chemicals differ from those of adults [1]. The activity of the immune system and of cell repair mechanisms of children is generally less developed [4,5]. Available data suggest that induction of genetic damage early in life in combination with an increased susceptibility of children exposed to environmental genotoxicants may increase risk of carcinogenesis in later life [6].

In molecular epidemiology, biomarkers are used to estimate exposures and (health) effects of environmental contaminants. This field of research would benefit from the development of new biomarkers or biomonitoring methods that would give more generic and mechanism-based information on several health effects than the currently available biomarkers, which are generally applicable only for one toxicological response or one class of compounds. Since the response of cells to environmental stress or xenobiotic exposure involves altered gene expression, it is thought that biomarkers based on gene expression analysis might be a promising new tool to fulfill these needs. The new field of toxicogenomics, and the new techniques that are emerging along, have provided the opportunity to investigate the expression of a large number of genes simultaneously, thereby obtaining information on a genome-wide response to *e.g.* environmental exposures. Toxicogenomics thus holds the platform for development of such new biomarkers [7] and has already been applied to exposed adult populations [8-11].

We have now performed an exploratory study on gene expression responses among children environmentally exposed to air pollutants. The children under study reside in the area of Teplice, a mining district situated in the northwest of the Czech Republic, where air pollution has been found to pose serious adverse health effects to humans [12]. In inhabitants of this region, the elevated concentrations of air pollutants have been demonstrated to result in measurable uptake, metabolism, and excretion of genotoxic organic compounds as well as in increased blood concentrations of toxic metals and in DNA damage [13]. Specifically in winter time, frequent periods of increased concentrations of air pollution are observed in this area, caused by the geographical location in a valley surrounded by mountains [14].

The objective of the present study was to analyze gene expression in peripheral blood from 23 children inhabiting the area of Teplice in the Czech Republic. For comparison, 24 children living in the area of Prachatice, representing a less-polluted rural area, were enrolled. Using genome-wide oligonucleotide microarrays, we investigated differential gene expression in the blood of Teplice children in comparison to children

from Prachatice which represents a rural and therefore less polluted area. As an additional approach, individual gene expressions were correlated with individual peripheral blood lymphocyte micronuclei frequencies, in order to investigate the linkage of gene expressions with an established biomarker of effect that is representative for increased genotoxic risk [15]. Micronuclei (MN) frequencies may thus serve as a phenotypic anchor for genotoxicity. Significantly affected biological pathways and processes will be reported.

Materials and methods

Study population

The study subjects were selected from the Teplice area in the north and from the rural area of Prachatice in the south of the Czech Republic. Whole blood was available from a total of 47 children; 23 from the Teplice area and 24 from the Prachatice area. The composition of the two populations was similar with respect to gender and age distribution (Table 1). Samples of blood from the children were drawn with the informed consent and permission of their parents. The study protocol was approved by the local ethical committee at the Institute of Experimental Medicine AS CR in Prague. The characterization of the air pollution in both regions is shown in Table 2 (unpublished data from R. Sram).

Table 1. *Characteristics of the study population*

	N subjects	Gender M / F	Age (mean $\pm$ se)	Age range
<u>Prachatice</u>				
Younger children	12	7 / 5	6.2 $\pm$ 0.5	5 - 7
Older children	12	6 / 6	9.4 $\pm$ 1.2	7 - 11
Total	24	13 / 11	7.8 $\pm$ 1.9	5 - 11
<u>Teplice</u>				
Younger children	12	6 / 6	6.5 $\pm$ 0.5	6 - 7
Older children	11	5 / 6	9.3 $\pm$ 1.2	7 - 11
Total	23	11 / 12	7.9 $\pm$ 1.7	6 - 11

Table 2. Average concentrations of air pollutants measured at the stationary monitoring stations in Prachatice (PR) and Teplice (TP) during the winter months preceding the blood sampling

	SO ₂	NO _x	NO ₂	NO	CO	PM2.5 ¹	PM10 ²	PAHs ^{2,3}	c- PAHs ^{2,4}
<u>3 months</u>									
PR	10.63	25.48	18.25	4.44	526.93	32.27	38.16	22.37	9.87
TP	17.47	66.05	35.85	19.69	662.31	30.11	37.22	88.90	18.31
<u>6 months</u>									
PR	12.70	25.25	18.08	4.40	607.44	36.15	41.44	19.10	8.32
TP	14.67	74.30	36.63	24.65	525.38	30.19	37.48	75.13	15.79

¹PM concentrations in µg/m³ air, average based on 24h mean concentrations

²PAH concentrations in ng/m³ air, average based on 24h mean concentrations

³Total of all PAHs measured

⁴Total of eight carcinogenic PAHs: benzo[a]pyrene, chrysene, benz[a]anthracene, benzo[b]fluoranthene, benzo[k]fluoranthene, dibenz[ah]anthracene, indeno[1,2,3-cd]pyrene and benzo[ghi]perylene.

Collection and processing of blood

Blood samples were collected in March 2004 in Prachatice and in Teplice. One older child from Teplice was excluded from blood collection due to ongoing health complications and consequent treatment with antibiotics. Blood was drawn into one PAXgene Blood RNA tube (PreAnalytix, Qiagen, Hilden, Germany) per child, which allowed for instant conservation of the RNA. The blood was drawn by qualified paediatric nurses at local health centres in the Czech Republic and transported to Prague at 17-22°C where the tubes were frozen to -80°C on the same day of blood collection. Blood samples were transported and stored at -80°C until RNA isolation.

RNA isolation

Total RNA was isolated from the PAXgene Blood RNA tubes using the PAXgene Blood RNA kit (PreAnalytix) according to the manufacturer's protocol with minor modifications. After total RNA isolation, the RNA was re-purified in order to remove possible genomic DNA by a DNase treatment (RNase free DNase set) and concentrated using the MinElute system (Qiagen, Hilden, Germany). Additionally, the RNA was purified from genomic DNA by a DNase treatment and concentrated using the MinElute system (Qiagen, Hilden, Germany). Final integrity of total RNA was verified using the BioAnalyzer (Agilent, Palo Alto, CA, USA) and concentrations and purity (A_{260}/A_{280} ratios) were determined with the Nanodrop (NanoDrop Technologies, Wilmington, USA).

Microarray analyses of gene expression

One microgram of total RNA was used for labelling with fluorophores using the Low Input RNA Linear Amplification method (Agilent Technologies, Palo Alto, CA, USA), conducted according to the producer's instructions. One microgram of total RNA from each child from Prachatice was used to generate a common reference pool. Briefly, first the total RNA was reverse transcribed into cDNA using T7-promoter primer and MMLV reverse

transcriptase. The cDNA was transcribed into cRNA, during which it was fluorescently labelled by incorporation of Cyanine5-CTP or Cyanine 3-CTP (Perkin Elmer, Boston, MA, USA). The samples from the children from both areas were labelled with Cy5, whereas the common reference samples were labelled with Cy3. After purification, using the RNeasy Mini kit (Qiagen), the probes were prepared for hybridization to Agilent Human 22k oligonucleotide microarrays. The incorporation efficiency of fluorophore into the cRNA was determined spectrophotometrically and only samples showing a dye-incorporation rate of >7 pmol/ μ g cRNA were considered suitable for further processing. Each array was hybridized with a mix of equal amounts of cRNA from a child and the reference RNA sample, which generated 47 microarrays. The arrays were hybridized for 17 hours at 60° C in a hybridisation oven. After washing according to the manufacturer's protocol, the slides were dipped in Drying and Stabilization Buffer (Agilent) (2:1 diluted with acetonitrile) to dry and protect them from environmental ozone.

Microarray data analysis

The arrays were scanned within 1 hour after hybridization was finished, using a ScanArrayExpress scanner (Perkin Elmer, Boston, MA, USA), which generated separate Cy5 and Cy3 tif-images. Laser power was set to 100% and the photomultiplier gain to 60% for Cy3 and 55% for Cy5. The image files were uploaded pair wise into ImaGene (BioDiscovery, Marina del Rey, CA, USA) to convert the images into raw data files. These files were then uploaded into the data analysis program GeneSight (BioDiscovery). Data transformations were applied in the following order: subtraction of local background, omission of bad spots (defined automatically by ImaGene based on the shape and signal intensity of the spot and background) and negative spots (automatically defined by ImaGene as having a lower signal than background), elimination of spots with signal intensities <20 pixels in both the Cy5 and Cy3 channel, log (base 2) transformation, locally weighted scatter plot smoothing normalization (LOWESS; global with a 0.2 smoothing factor) and calculation of the difference of the child's sample with the common reference sample.

These expression differences relative to the reference sample of all 47 children were used for the subsequent (statistical) analyses. In order to select the genes that are differentially expressed between both populations and genes that correlate with micronuclei levels, the web-based automated POMELO tool from Gene Expression Pattern Analysis Suite (GEPAS v1.1; <http://gepas.bioinfo.cnio.es/index.html>) was used. First, genes with $>70\%$ missing values were filtered out, followed by imputation of remaining missing values through K-nearest neighbour computation ($K=10$). In the statistical tests 50.000 permutations were applied, in order to reduce the number of false positive calls (a well known problem when performing multiple tests on multiple testing of thousands of genes). Within GEPAS, a t-test was applied to select the genes that are significantly differentially expressed between Teplice and Prachatic. Also within GEPAS, Pearson linear regression analysis was used to investigate significant associations (correlations) of individual gene expressions with individual micronuclei frequencies. For all statistical analyses applied, p-values <0.05 were considered significant.

Principal component analysis was performed in GeneSight, as well as a hierarchical clustering analysis (average linking, Euclidean clustering), in order to visualize the

distinction of exposed and non-exposed children based on a selected set of significantly modulated genes.

Functionalities of genes of interest were analyzed using the online tool Expression Analysis Systematic Explorer (EASE, <http://apps1.niaid.nih.gov/David>) [16] to investigate biological enrichment of certain pathways or processes. This was done for the sets of genes with p-values of the difference between the two regions <0.01 and <0.05. Biological processes showing EASE scores <0.05 are reported significant.

Real-time PCR analysis of gene expression

Expressions of a number of selected genes were verified by means of SYBR Green I real-time PCR (BioRad Laboratories, Hercules, USA), using aliquots of the cDNA that was synthesized for the microarray hybridizations as described above for each child. In each run, a dilution series of the common reference pool was included as well as no template controls to estimate PCR efficiency. Real-time PCR reactions were performed on the BioRad MyiQ iCycler Single Color real-time PCR detection system using iQ™ SYBR® Green Supermix, containing iTaq Polymerase, dNTPs, SYBR Green I and buffers (BioRad). Reactions were initiated for 3 min at 95°C to activate Taq polymerase, followed by 40 cycles of 15 seconds at 95°C and 45 seconds at 60°C. Melting curve analysis was performed starting at 60°C with stepwise temperature elevations of 0.5°C every 10 seconds to check for nonspecific products. Two stably expressed genes were included as internal controls; beta-actin (*ACTB*) and cyclophilin A (*PPIA*). All reactions were performed in duplicate.

Primer sequences were as follows:

CXCL1

5'-CCACTGCGCCCAAACC-3' (forward) and 5'-GCAGGATTGAGGCAAGCTTT-3' (reverse),

TIGD3

5'-GTGCTGGAAGCTCTGGATGAG-3' (forward) and 5'-TTGCAGATGCGCGAGATCT-3' (reverse)

DGAT2

5'-GCACAGAGGCCACAGAAGTG-3' (forward) and 5'-CCCTCAACACAGGCATTGCG-3' (reverse)

PINK1

5'-AGCAGTCACTTACAGAAAATCCAAGA-3' (forward) and 5'-GGTGAAGCGCGGAGAA-3' (reverse),

FXN

5'-CAACAAGCAGACGCCAAACA-3' (forward) and 5'-ACGCCGTCGTGGGAGTAC-3' (reverse),

ACTB

5'-CCTGGCACCCAGCACAAAT-3' (forward) and 5'-GCCGATCCACACGGAGTACT-3' (reverse),

PPIA

5'-TTCCTGCTTTCACAGAATTATTCC-3' (forward) and 5'-GCCACCAGTGCCATTATGG-3' (reverse).

Ct values per child and per gene were normalized by subtraction of the mean Ct values of the internal control genes. Subsequently, $\Delta\Delta Ct$ values were calculated per child and per gene relative to the normalized pooled reference sample. This equals the following formula:

$$\Delta\Delta Ct = \text{pooled reference } (Ct_{\text{gene}} - Ct_{\text{mean } ACTB-PPIA}) - \text{child } (Ct_{\text{gene}} - Ct_{\text{mean } ACTB-PPIA})$$

This returned one expression value per gene per individual child. All gene expression data are reported as $\Delta\Delta Ct$ values and are expressed on log (base = 2) scale.

Analysis of micronuclei frequencies

The cytokinesis-block technique was used to analyze micronuclei (MN) frequencies [17]. Each lymphocyte culture was incubated for 72 hours at 37°C. Cell divisions were arrested after 44 hours using cytochalasin B. The slides were coded and examined at 400x magnification via a light microscope. For each individual, a total of 2000 binucleated cells were evaluated for the presence of MN. Criteria for identifying viable cells for scoring of MN and MN were previously described by the HUMN project [18]. A detailed description of MN frequency distribution in the children will be reported separately by Pedersen *et al.*

Results

Concentrations of air pollutants in ambient air are depicted in Table 2, showing average concentrations over 3 months (Dec-Feb) and 6 months (Sep-Feb) preceding the blood collection. For each pollutant, higher levels were observed in the Teplice area than in the Prachatice area, except for PM_{2.5} and 10 concentrations and for CO measured over 6 months. Concentrations of polycyclic aromatic hydrocarbons (PAH) and the carcinogenic fraction of all measured PAH, were respectively four-fold and nearly two-fold higher in the Teplice region than in the Prachatice region.

Microarray analyses of gene expression

The isolated total RNA was found to be of high purity and quality with mean A_{260}/A_{280} ratios of 2.0 (range 1.9 – 2.1) and a mean RNA Integrity Number (RIN) of 8.3 (range 7.5 - 9.1).

Microarray analysis of gene expression yielded considerable numbers of genes which appeared significantly differentially expressed in the Teplice group as compared to the Prachatice group; 1727 genes showed an unadjusted p-value of <0.05 (Table 3). Of these genes, 1001 were up regulated and 726 were down regulated in the children from Teplice relative to the children from the Prachatice area. After performance of the permutation t-test with 50.000 permutations, 2 genes remained with adjusted p-values <0.05, *i.e.* chromosome 15 open reading frame 16 (*C15orf16*) and dual specificity phosphatase-like 15 (*DUSP15*). Table 4 shows the expression differences of the 26 genes that most significantly differed between children from the two regions.

Table 3. Overview of the numbers of genes significantly differentially expressed in children from Teplice as compared to children from Prachatice

p-value (2-tailed significance)	# genes differentially expressed in Teplice versus Prachatice children		
	total	up-regulated	down-regulated
<0.0001	26	25	1
<0.001	95	81	14
<0.01	487	315	172
<0.05	1727	1001	726

Table 4. Expression changes of the 26 genes that most significantly differed between children from Teplice (TP) and Prachatice (PR) ($p < 0.0001$)

Gene	GenBank Accession #	p-value	Effect (difference TP-PR)
<i>C15orf16</i>	NM_130901	4.97×10^{-7}	0.42
<i>DUSP15</i>	NM_080611	2.59×10^{-6}	0.39
<i>GPR20</i>	NM_005293	4.93×10^{-6}	0.31
<i>FHKL18</i>	NM_004118	1.25×10^{-5}	0.38
<i>Unknown</i>	A_23_42991	1.60×10^{-5}	0.36
<i>NEUROG3</i>	NM_020999	1.61×10^{-5}	0.45
<i>TTS</i>	NM_003312	1.66×10^{-5}	0.35
<i>PINK1</i>	NM_032409	2.32×10^{-5}	0.34
<i>COX6A2</i>	NM_005205	2.32×10^{-5}	0.37
<i>HOWB8</i>	NM_024016	2.69×10^{-5}	0.32
<i>FXN</i>	U43747	2.85×10^{-5}	0.28
<i>MPST</i>	NM_021126	3.72×10^{-5}	0.29
<i>PCSK1N</i>	NM_013271	3.88×10^{-5}	0.32
<i>EMCN</i>	AL133118	4.19×10^{-5}	-0.39
<i>Unknown</i>	THC1881025	4.55×10^{-5}	0.31
<i>Unknown</i>	A_23_P332782	4.90×10^{-5}	0.35
<i>GPSM3</i>	NM_022107	4.96×10^{-5}	0.29
<i>SRGAP2</i>	NM_014850	5.56×10^{-5}	0.29
<i>CCS</i>	NM_005125	6.16×10^{-5}	0.19
<i>PP2447</i>	NM_025204	6.33×10^{-5}	0.35
<i>BBC3</i>	NM_014417	6.69×10^{-5}	0.32
<i>CXCL1</i>	NM_001511	6.70×10^{-5}	0.52
<i>GPR153</i>	NM_207370	7.19×10^{-5}	0.33
<i>TIP39</i>	NM_178449	8.13×10^{-5}	0.27
<i>GPR78</i>	NM_080819	8.45×10^{-5}	0.36
<i>HISTH1C</i>	NM_005319	9.21×10^{-5}	0.33

The hierarchical clustering analysis based on the set of genes differentially expressed at $p < 0.001$ (two-sided unpaired t-test assuming equal variances) returned a quite distinct separation of exposed and non-exposed individuals (Figure 1), showing two relatively large clusters for exposed and non-exposed individuals, next to a few smaller clusters. The principal component analysis (PCA) based on the same gene set confirmed this observation as it showed two clouds that are mainly distinct but overlap in a small number of cases (Figure 2).

Figure 1. Hierarchical clustering (average linkage, euclidean clustering) based on the set of genes that were differentially expressed in children from Teplice (black condition lines) as compared to children from Prachatice (grey condition lines) at $p < 0.001$ ($n = 95$, see also Table 3). Individuals are on the x-axis and genes on the y-axis. Grey: up-regulation and black: down-regulation (difference of child's sample and the common reference sample).

Figure 2. Principal Component Analysis based on the set of genes that were differentially expressed between Teplice (black squares) and Prachatice (grey squares) at $p < 0.001$ ($n = 95$).

Analysis of the list of genes that were differentially expressed at $p < 0.05$ for overrepresentation of biological processes or molecular functions using the Expression Analysis Systematic Explorer (EASE), the most significantly affected biological processes were nucleosome assembly and chromatin assembly/disassembly (EASE scores $1 \cdot 10^{-3}$ and $2 \cdot 10^{-3}$ respectively). For the list of genes differentially expressed at $p < 0.01$, the same processes were found most significantly affected at EASE scores of $1 \cdot 10^{-3}$ and $4 \cdot 10^{-3}$ respectively. The enrichment of these processes in terms of a decrease in EASE scores became more pronounced when analyzing the up and down regulated genes separately. In the portion of only up regulated genes at $p < 0.01$ ($n=315$), these two processes showed EASE scores of $2 \cdot 10^{-4}$ and $3 \cdot 10^{-4}$ respectively. After analyzing genes up regulated at $p < 0.05$ ($n=1001$) the EASE scores decreased to $1 \cdot 10^{-5}$ and $2 \cdot 10^{-5}$. In contrast, in the set of genes that were down regulated in the children from Teplice in comparison with Prachatice, only at $p < 0.05$ significantly affected processes were found; nucleic acid metabolism (EASE score $1 \cdot 10^{-5}$) and RNA splicing and processing (EASE scores $1 \cdot 10^{-4}$) being the most significant ones. Next to nucleosome assembly and chromatin assembly/disassembly, other processes were found significantly affected by it at higher EASE scores but still at a significance level of $p < 0.05$ (Table 5A). Amongst these are e.g. microtubule based movement, M-phase associated processes, muscle development, immune response and vitamin metabolism.

Analysis of micronuclei frequencies

A higher average frequency of MN was found in the PBL of the children from the Teplice area, as compared to the Prachatice children (mean $\pm$ se 8.0 ± 3.2 MN % in children from the Teplice area versus 5.8 ± 3.4 MN % in children from the Prachatice area; $p=0.023$, unpaired 2-sided t-test, assuming equal variances).

The linear regression analysis between individual gene expressions and individual micronuclei frequencies returned significant associations for 1452 gene expressions, of which 705 were negatively and 747 were positively correlated with the frequency of micronuclei. The correlation coefficients ranged from 0.26 to 0.58 (positively associations) and from -0.27 to -0.56 (negative associations). Analysis of enrichment of biological processes or molecular functions within this set of genes revealed (m)RNA metabolism to be the most significantly affected biological process (EASE score 0.001, Table 5B). In the subset of genes that positively correlated with MN frequencies, the most affected process was cell communication whereas gene expressions negatively associated with micronuclei frequency, were related mainly to RNA processing.

Table 5. Overview of a selection of significantly overrepresented biological processes according to EASE analyses, in the sets of genes, that significantly differed in expression in children from Teplice compared to children from Prachatice (A) and of the gene expressions that significantly correlated with the individual micronuclei frequencies (B).

A		
Significance	Regulation	Biological process (EASE score)
p<0.01	Up	Nucleosome assembly (0.0002) Chromatin assembly/disassembly (0.0004) Microtubule based movement (0.006) DNA packaging (0.014) Muscle development (0.046)
	Down	-
	All	Nucleosome assembly (0.002) Chromatin assembly/disassembly (0.004) Microtubule based movement (0.021) DNA packaging (0.039)
p<0.05	Up	Nucleosome assembly (0.00001) Chromatin assembly/disassembly (0.00002) Cell communication (0.006) Innate immune response (0.015) Inflammatory response (0.016) Circulation (0.031) Muscle development (0.039)
	Down	Nucleobase, nucleoside, nucleotide and nucleic acid metabolism (0.00001) RNA splicing (0.0001) M-phase (0.0008) Nuclear division (0.001) Transcription (0.006) Response to DNA damage stimulus (0.014)
	All	Nucleosome assembly (0.001) Chromatin assembly/disassembly (0.002) Circulation (0.012) M phase of mitotic cell cycle (0.022) Vitamin metabolism (0.033) Nuclear division (0.048)
B		
Association with MN frequencies		Biological process (EASE score)
Positive (n=748)		Cell communication (0.00006) G-protein coupled receptor protein signalling pathway (0.002) Signal transduction (0.008) Response to chemical substance (0.015) Cell differentiation (0.023) MAPKKK cascade (0.040) Innate immune response (0.049)
Negative (n=706)		RNA processing (2×10^{-9}) Protein biosynthesis (0.001) Exogenous antigen processing (0.001) Cell cycle checkpoint (0.034) Mitotic cell cycle (0.043)
All (n=1454)		RNA metabolism (0.001) Exogenous antigen processing (0.015) Cell differentiation (0.019) Vasoconstriction (0.033)

Real-time PCR analyses of gene expression

Real-time PCR analysis was performed on a selection of genes based on the microarray results. Genes were chosen on the basis of a combination of two criteria: discrimination of children from the two regions most significantly (*i.e.* presenting the lowest *p*-values of the difference between the two regions) and showing strong and significant correlation with micronuclei frequencies. Data analyses returned eight genes that responded best to these criteria; chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha) (*CXCL1*), PTEN-induced putative kinase 1 (*PINK1*), diacylglycerol O-acyltransferase homolog 2 (mouse) (*DGAT2*), tigger transposable element derived 3 (*TIGD3*), frataxin (*FXN*), neurogenin 3 (*NEUROG3*), chromosome 15 open reading frame 16 (*C15orf16*) and forkhead-like 18 (*FKHL18*).

Table 6. Real-time PCR verification results as compared with microarray results for the selected gene expressions. Numbers in brackets represent *p*-values. Difference means difference between the Teplice children and the Prachatice children. Correlation values represent Pearson correlation coefficients (*R*).

Gene	Microarray		Real-time PCR	
	Difference	Correlation	Difference	Correlation
<i>CXCL1</i>	0.52 (0.00007)	0.37 (0.011)	0.60 (0.0009)	0.36 (0.014)
<i>PINK1</i>	0.34 (0.00002)	0.30 (0.038)	0.37 (0.009)	0.16 (0.27)
<i>DGAT2</i>	0.43 (0.0002)	0.32 (0.027)	0.78 (0.005)	0.25 (0.09)
<i>TIGD3</i>	0.033 (0.0002)	0.41 (0.004)	0.52 (0.008)	0.39 (0.008)

Real-time PCR verification of the microarray results showed the *FKHL18*, *C15orf16* and *NEUROG3* genes to have too low abundance to be reliably measured, showing threshold cycle values above 30. The *CXCL1*, *PINK1*, *DGAT2* and *TIGD3* gene expressions as measured with real-time PCR confirmed the microarray results (Table 6). Expression of the *FXN* gene did not confirm the microarray-based findings and is therefore likely to be a false-positive result.

Discussion

This study presents the first investigation of genome-wide differential gene expression in a human population that is exposed to air pollution at daily life conditions, the first report on monitoring transcriptomic responses to air pollution in children and the first attempt to correlate gene expressions with a biomarker of genotoxicity; micronuclei frequencies in peripheral lymphocytes *in vivo*. Differential gene expression was investigated in the blood of children from the Czech Republic; a group of children residing in the area of Teplice, which is characterized by relatively high levels of air pollution and a group of children inhabiting the area of Prachatice, with lower concentrations of air pollutants (Table 2).

Changes related to the exposure

Using human oligonucleotide microarrays containing 22k gene sequences, a global analysis of genome-wide gene expression changes was performed which delivered considerable numbers of genes significantly differentially expressed in the Teplice group as compared to the Prachatice group. *C15orf16* and *DUSP15* are the two genes that most significantly differ in expression between the Teplice children and the Prachatice children (p-values 4.97×10^{-7} and 2.59×10^{-6} respectively). *C15orf16* is also known as *OTUD7* (OTU domain containing 7) and its functions are associated with DNA binding, cysteine-type peptidase activity, metal (zinc) ion binding and the ubiquitin cycle (according to Gene Ontology at <http://www.ncbi.nlm.nih.gov>). The exact function of this gene is unknown. *DUSP15* encodes a protein that belongs to the protein-tyrosine phosphatase family, having both protein-tyrosine phosphatase activity and serine/threonine-specific phosphatase activity. For both these genes it is not clear at this moment exactly how they biologically relate to exposure air pollution. However, since they discriminate children from both regions with such high significance, elucidation of their function could be subject to further investigation.

Within the total set of significantly differentially expressed genes, the main enriched biological process according to Expression Analysis Systematic Explorer (EASE) analysis was nucleosome assembly (EASE score 0.001, Table 5A). Analysis of the proportion of only up regulated genes ($n=1001$) returned a substantially better EASE score of 1×10^{-5} for this biological process. For the subset of genes that were down regulated ($n=726$), EASE showed nucleobase, nucleoside, nucleotide and nucleic acid metabolism and RNA splicing and processing to be the main affected biological processes (EASE scores 1×10^{-5} and 1×10^{-4} respectively). The nucleosome assembly pathway encompasses mainly genes that encode histones. The modulation of this process might indicate nucleosome assembly being deregulated, and suggests an impact on the condensing of DNA by environmental air pollution, since the nucleosome is the primary structural unit of chromatin. Chromatin is a pivotal player in the regulation of gene expression, chromosome replication and cell cycle progression. It is also said to maintain genome integrity by marshalling access to the DNA template [20]. Other related biological processes that were found to be modulated were microtubule based movement, DNA packaging and nuclear division. In addition to these findings, other more generic biological processes were found to be modulated by exposure to air pollution including processes relevant to children's health and development, like muscle development, innate immune response and vitamin metabolism.

The results of the hierarchical clustering and the principal component analysis, based on the set of genes showing differential expression at $p < 0.001$ (unpaired 2-tailed t-test assuming equal variances, Figures 1 and 2), also demonstrate discriminating gene expression profiles between inhabitants of the two regions. With the hierarchical clustering (Figure 1), potential effects of age, gender and genetic background have been studied. Effects of age, gender or family were not found, *i.e.* children did not cluster together according to age or gender and siblings did not cluster together as well.

Changes associated with micronuclei

Individual gene expressions significantly correlated with individual micronuclei frequencies which represent an established phenotypic endpoint of genotoxicity. Average micronuclei frequencies appeared to differ significantly between the Teplice and the Prachatice group (p-value 0.023). After linear regression analysis using GEPAS, a total of 1452 genes were observed that significantly correlated with individual micronuclei frequencies. These genes may be hypothesized to possess effect marker potential because of their phenotypical anchoring with this biomarker that has been demonstrated to be predictive of genotoxic risk. The main affected process observed in this total set of genes was RNA metabolism (EASE score 0.001, Table 5B), which indicates that the frequency of micronuclei in the cell has an impact on the processing of RNA. Although the average micronuclei frequencies differed significantly between inhabitants of the two regions (p=0.023), also a clear overlap existed between both groups, which might explain the differences in affected biological processes observed in the whole set of significantly differentially expressed genes versus the subset of gene expressions that correlated with micronuclei frequencies.

For verification of gene expressions, based on the microarray results, with an independent assay, genes were selected that answered to the criteria of significantly discriminating children from the two regions and also showing a significant association of the individual gene expression with individual MN frequencies. Real-time PCR analyses confirmed the differential expressions of the *CXCL1*, *PINK1*, *DGAT2* and *TIGD3* genes between children from Teplice and Prachatice regions (Table 6). Also, the significant correlations of the *CXCL1*, *DGAT2* and *TIGD3* expressions with MN frequencies were reproduced with real-time PCR. The association of *PINK1* expression with MN frequency returned a weaker and non-significant association as analyzed with real-time PCR. The main or most important functions of these genes that might be relevant for environmentally exposed children, according to Gene Ontology at <http://www.ncbi.nlm.nih.gov/Gene>, are as follows. *CXCL1* is a family member of the chemokines, which regulate cell trafficking of various types of leukocytes through interactions with a subset of 7-transmembrane, G protein-coupled receptors. Chemokines also play an important role in the development, homeostasis, and function of the immune system, and they have effects on cells of the central nervous system. *PINK1* over-expression has been reported to be involved in stress response, in particular in the protection of neurons against stress-induced mitochondrial dysfunction and apoptosis [21]. *DGAT2* is involved in the synthesis of triglycerides and obesity [22]. Finally, the protein encoded by the *TIGD3* gene belongs to the superfamily of DNA-mediated transposons in humans, however the exact function of this gene is not known.

It has been suggested that the induction of differential gene expression as measured with microarrays is not an equally sensitive marker of genotoxicity as the classical genotoxicity endpoints such as micronuclei are [23]. Dickinson *et al.* found in mouse lymphoma cells that *in vitro* higher exposures (of cisplatin) were required to induce robust gene expression changes than to induce micronuclei [24]. Forrest *et al.* have reported recently on monitoring gene expression in a human population that was occupationally exposed to benzene [9] and found genes that potentially mark benzene exposure. They however did not associate their gene expressions with a marker of a

genotoxic endpoint. In the present human *in vivo* study we found a considerable set of gene expressions that significantly differed at a relatively low dose exposure to air pollutants, when compared to other studies investigating *e.g.* occupational exposures that are generally higher (Table 2), as well as a substantial set of genes that correlated significantly with micronuclei frequencies. In the children from the exposed area of Teplice a 1.38 fold increase of micronuclei frequency was measured as compared to children from the Prachatice area (mean Teplice 8.0 MN ‰ versus mean Prachatice 5.8 MN ‰), and we observed a 1.24 to 1.34 fold difference in gene expression of the five genes that were most significantly modulated. The expression differences of these genes however, segregate the two groups at considerable lower p-values than the micronuclei frequencies do while showing less overlap between exposed and non-exposed individuals. The relatively small differences in gene expression that were observed *in vivo* as compared to previously reported *in vitro* studies on blood cells [25-27], are not uncommon in *in vivo* blood gene expression profiling [11], and therefore do not present analytical problems. Gene expression analysis may thus very well yield a promising indicator of genotoxicity in exposed humans that is not considered less sensitive than classical markers.

Because of the apparently large numbers of genes that significantly differed in expression in the exposed children from the Teplice area, as compared to the children from the control area of Prachatice, it appears feasible to generate discriminative profiles of gene expression modulations and of biological processes at rather low differences in exposure levels of environmental carcinogens. Based on the results of this study among children exposed to ambient air pollution, we suggest that transcriptomic analysis represents a promising tool for monitoring adverse health effects due to environmental exposure. It is therefore of relevance to monitor more exposed populations for transcriptomic responses to environmental contaminants.

Acknowledgements

We greatly acknowledge the ChildrenGenoNetwork and its participants for their efforts. We also thank the pediatricians from Vimperk, Krupka and Teplice for blood collection. This study was supported by the ChildrenGenoNetwork (QLRT-2001-02198) and by the Flanders Environment and Health Study.

References

1. Suk, W.A., Murray, K. and Avakian, M.D. (2003). Environmental hazards to children's health in the modern world. *Mutat Res* 544, 235-242.
2. Landrigan, P.J., Kimmel, C.A., Correa, A. and Eskenazi, B. (2004). Children's health and the environment: public health issues and challenges for risk assessment. *Environ Health Perspect* 112, 257-265.
3. Neri, M., Bonassi, S., Knudsen, L.E., Sram, R.J., Holland, N., Ugolini, D. and Merlo, D.F. (2006). Children's exposure to environmental pollutants and biomarkers of genetic damage. I. Overview and critical issues. *Mutat Res* 612, 1-13.
4. Chance, G.W. and Harmsen, E. (1998). Children are different: environmental contaminants and children's health. *Can J Public Health* 89 Suppl 1, S9-13, S10-15.
5. Rossner, P., Bavorova, H., Ocadlikova, D., Svandova, E. and Sram, R.J. (2002). Chromosomal aberrations in peripheral lymphocytes of children as biomarkers of environmental exposure and life style. *Toxicol Lett* 134, 79-85.
6. Wild, C.P. and Kleinjans, J. (2003). Children and increased susceptibility to environmental carcinogens: evidence or empathy? *Cancer Epidemiol Biomarkers Prev* 12, 1389-1394.
7. Toraason, M., Albertini, R., Bayard, S., Bigbee, W., Blair, A., Boffetta, P., Bonassi, S., Chanock, S., Christiani, D., Eastmond, D., Hanash, S., Henry, C., Kadlubar, F., Mirer, F., Nebert, D., Rapport, S., Rest, K., Rothman, N., Ruder, A., Savage, R., Schulte, P., Siemiatycki, J., Shields, P., Smith, M., Tolbert, P., Vermeulen, R., Vineis, P., Wacholder, S., Ward, E., Waters, M. and Weston, A. (2004). Applying new biotechnologies to the study of occupational cancer--a workshop summary. *Environ. Health Perspect* 112, 413-416.
8. Lampe, J.W., Stepaniants, S.B., Mao, M., Radich, J.P., Dai, H., Linsley, P.S., Friend, S.H. and Potter, J.D. (2004). Signatures of environmental exposures using peripheral leukocyte gene expression: tobacco smoke. *Cancer Epidemiol. Biomarkers Prev* 13, 445-453.
9. Forrest, M.S., Lan, Q., Hubbard, A.E., Zhang, L., Vermeulen, R., Zhao, X., Li, G., Wu, Y.Y., Shen, M., Yin, S., Chanock, S.J., Rothman, N. and Smith, M.T. (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.
10. 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.
11. Wang, Z., Neuburg, D., Li, C., Su, L., Kim, J.Y., Chen, J.C. and Christiani, D.C. (2005). Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. *Environ Health Perspect* 113, 233-241.
12. Sram, R. Teplice Program: Studies on the impact of air pollution on human health (1991-1999), in: R. Sram (Ed.), *Teplice Program: Impact of Air Pollution on Human Health*, Academia, Prague, 2001, pp19-29.
13. Sram, R.J., Benes, I., Binkova, B., Dejmek, J., Horstman, D., Kotesovec, F., Otto, D., Perreault, S.D., Rubes, J., Selevan, S.G., Skalik, I., Stevens, R.K. and Lewtas, J. (1996). Teplice program - the impact of air pollution on human health. *Environ Health Perspect* 104 Suppl 4, 699-714.
14. Binkova, B., Vesely, D., Vesela, D., Jelinek, R. and Sram, R.J. (1999). Genotoxicity and embryotoxicity of urban air particulate matter collected during winter and summer period in two different districts of the Czech Republic. *Mutat Res* 440, 45-58.
15. Fenech, M. (1993). The cytokinesis-block micronucleus technique: a detailed description of the method and its application to genotoxicity studies in human populations. *Mutat Res* 285, 35-44.
16. Hosack, D.A., Dennis, G., Jr., Sherman, B.T., Lane, H.C. and Lempicki, R.A. (2003). Identifying biological themes within lists of genes with EASE. *Genome Biol* 4, R70.
17. Fenech, M. and Morley, A.A. (1985). Measurement of micronuclei in lymphocytes. *Mutat Res* 147, 29-36.
18. Fenech, M., Chang, W.P., Kirsch-Volders, M., Holland, N., Bonassi, S. and Zeiger, E. (2003). HUMN project: detailed description of the scoring criteria for the cytokinesis-block micronucleus assay using isolated human lymphocyte cultures. *Mutat Res* 534, 65-75.
20. Moggs, J.G. and Orphanides, G. (2004). The role of chromatin in molecular mechanisms of toxicity. *Toxicol Sci* 80, 218-224.
21. Valente, E.M., Abou-Sleiman, P.M., Caputo, V., Muqit, M.M., Harvey, K., Gispert, S., Ali, Z., Del Turco, D., Bentivoglio, A.R., Healy, D.G., Albanese, A., Nussbaum, R., Gonzalez-Maldonado, R., Deller, T., Salvi, S., Cortelli, P., Gilks, W.P., Latchman, D.S., Harvey, R.J., Dallapiccola, B., Auburger, G. and Wood, N.W. (2004). Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science* 304, 1158-1160.

22. Suzuki, R., Tobe, K., Aoyama, M., Sakamoto, K., Ohsugi, M., Kamei, N., Nemoto, S., Inoue, A., Ito, Y., Uchida, S., Hara, K., Yamauchi, T., Kubota, N., Terauchi, Y. and Kadowaki, T. (2005). Expression of DGAT2 in white adipose tissue is regulated by central leptin action. *J Biol Chem* 280, 3331-3337.
23. Newton, R.K., Aardema, M. and Aubrecht, J. (2004). The utility of DNA microarrays for characterizing genotoxicity. *Environ Health Perspect* 112, 420-422.
24. Dickinson, D.A., Warnes, G.R., Quievryn, G., Messer, J., Zhitkovich, A., Rubitski, E. and Aubrecht, J. (2004). Differentiation of DNA reactive and non-reactive genotoxic mechanisms using gene expression profile analysis. *Mutat Res* 549, 29-41.
25. Amundson, S.A., Do, K.T., Shahab, S., Bittner, M., Meltzer, P., Trent, J. and Fornace, A.J., Jr. (2000). Identification of potential mRNA biomarkers in peripheral blood lymphocytes for human exposure to ionizing radiation. *Radiat. Res* 154, 342-346.
26. Ryder, M.I., Hyun, W., Loomer, P. and Haqq, C. (2004). Alteration of gene expression profiles of peripheral mononuclear blood cells by tobacco smoke: implications for periodontal diseases. *Oral Microbiol. Immunol* 19, 39-49.
27. van Leeuwen, D.M., Gottschalk, R.W., van Herwijnen, M.H., Moonen, E.J., Kleinjans, J.C. and van Delft, J.H. (2005). Differential gene expression in human peripheral blood mononuclear cells induced by cigarette smoke and its constituents. *Toxicol Sci* 86, 200-210.

Chapter 6

Biomonitoring of differential gene expressions in peripheral blood in selected environmentally exposed human populations in Flanders

D.M. van Leeuwen¹

R.W.H. Gottschalk¹

G. Schoeters²

J.C.S. Kleinjans¹

J.H.M. van Delft¹

¹ Department of Health Risk Analysis and Toxicology,
Maastricht University, Maastricht,
the Netherlands

² Flemish Institute for Technological Research,
Mol, Belgium

Abstract

A part of the incident cancer cases is known to be initiated and/or promoted by exposure to environmental chemicals with carcinogenic properties that occur in the human environment. In theory, that part of the incident cancer cases should thus be preventable. Molecular cancer epidemiology seeks to identify human cancer risks posed by exposure to environmental carcinogens by using biomarkers. Toxicogenomic technologies provide the opportunity to return potential biomarkers consisting of altered gene expression profiles. Based on two previous biomonitoring studies, eight genes were selected as potential biomarkers for carcinogenic exposure. The aim of the present study was to monitor the expression of these genes in a random sample of adults of the selected human populations in Flanders. The effect of the region of inhabitation on gene expression was investigated, whereby regions were selected based on (theoretical) different environmental burdens. Furthermore, associations of gene expressions with established biomarkers of early biological effect and tumor markers were investigated. Individual gene expressions of *CYP1B1*, *ATF4*, *MAPK14*, *SOD2*, *CXCL1*, *DGAT2*, *TIGD3* and *PINK1* were measured by means of real-time PCR using RNA extracted with the PAXgene system from peripheral blood cells of 398 individuals from Flanders. After correction for the confounding effect of tobacco smoking, gene expressions showed to differ significantly between two or more regions of inhabitation. In addition, significant correlations of the gene expressions with an early biological effect marker and tumor markers were observed. This study shows that gene expressions or a gene expression profile can segregate individuals according to the region they inhabit. Gene expression profiling therefore appears promising for application in human populations to analyze carcinogenic risks due to environmental exposures.

Introduction

Cancer is among the major causes of death in our society today [1]. Carcinogenesis, the process of the development of cancer, is known to be initiated and/or promoted by exposure to chemicals with carcinogenic properties that occur in the human environment, *e.g.* from ambient air, food or life style related factors such as cigarette smoke [2-5]. In theory, a part of the incident cancer cases should thus be preventable [6]. Molecular cancer epidemiology seeks to identify human cancer risks as posed by exposure to environmental carcinogens. To this end, biomarkers are investigated which enable risk analysis and assessment long before traceable, diagnostic health effects appear. Monitoring of these biomarkers frequently encompasses measurement of biological events that indicate the cancer risk but do not necessarily lead to the disease. These measurements may provide information that subsequently might serve for prevention purposes through policy measures by the government.

In monitoring environmental cancer risk among human populations over the last decades, several biological markers of internal carcinogen dose or their metabolites in blood or urine, and of early biological effects, *e.g.* DNA damage markers such as micro-nuclei (MN) and DNA strand breaks, have been applied [7]. Because of practical and ethical issues, biomonitoring relies on the use of surrogate target cells, *i.e.* cells that are assumed to at least partially reflect the changes in the target tissue and that are obtainable through minimally invasive procedure. The main surrogate target cells that are used in molecular epidemiology are blood cells.

Emerging new technologies that have given rise to the new field of toxicogenomics [8-10], such as microarrays and real-time PCR, have provided the opportunity to explore (changes in) the expression of large numbers of genes simultaneously and in a high-throughput manner [11-15]. In general, microarrays are applied for the analysis of differential gene expression for very large numbers of genes within a relatively small set of samples, *e.g.* in experimental studies, whereas real-time PCR is often applied for the analysis of a small set of genes within a large(r) set of samples, *e.g.* a human population. Next to the currently used biomarkers, alterations in the expression of genes are thought to occur as early events in the cell after exposure to, *e.g.* environmental toxicants and carcinogens.

Studies on environmental carcinogen-induced differential gene expression have been conducted in our laboratory previously. Besides experimental *in vitro* studies that were directed to obtaining gene expression profiles from model carcinogens (cigarette smoke exposure) in peripheral blood mononuclear cells (presented in **Chapter 2**) [16], a small human population of smoking-discordant monozygotic twin pairs has been investigated and genes were identified the expression of which significantly differed in smokers as compared to their non-smoking, genetically identical siblings (presented in **Chapter 3**) [17]. Also, a population study has been conducted among children inhabiting an air-polluted area in the Czech Republic. Genome-wide gene expressions in these children have been compared to those of children from an area with lower levels of air pollutants in the Czech Republic and have been correlated with MN frequencies (**Chapter 5**) [18]. These studies have identified environmentally induced gene expression changes, which

have served as a basis for the selection of genes for the large scale biomonitoring study presented in this chapter.

The aim of the present study was to monitor the expression of selected genes in the general human population and to correlate these to established biomarkers of early biological effect. In a sample of the general population, the expression levels of eight key genes, *CYP1B1*, *ATF4*, *MAPK14*, *SOD2*, *CXCL1*, *DGAT2*, *TIGD3* and *PINK1* [17,18] (Table 1), were measured in blood cells by means of real-time PCR. The present study was conducted as a part of the Flanders Environment and Health Study (FLEHS). FLEHS is a multi-academic consortium, instigated to research the impact of the environment on human health across the general Flanders population. Since the study participants were recruited from eight regions of inhabitation across Flanders, each with its specific environmental characteristics (Table 2), we examined gene expression for significant difference between groups of individuals inhabiting different regions. Also, gene expressions were investigated for correlation with available data on early biological effect markers (DNA strand breaks and micronuclei (MN) frequencies) and serum levels of several tumor markers (prostate-specific antigen (PSA), carcino-embryonic antigen (CEA) and p53).

Table 1. Overview of the genes monitored for their expression in blood cells

Gene name (abbreviation) ¹	GenBank Accession # ¹	Biological summary ¹
Cytochrome p450 1B1 (<i>CYP1B1</i>)	NM_000104	Member of the cytochrome P450 superfamily of enzymes; monooxygenases that catalyzes many reactions involved in drug metabolism. Metabolism of procarcinogens, e.g. polycyclic aromatic hydrocarbons (PAHs).
Activating transcription factor 4 (<i>ATF4</i>)	NM_001675	Encodes a transcription factor that belongs to a family of DNA-binding proteins, including the AP-1 family of transcription factors, cAMP-response element binding proteins (CREBs) and CREB-like proteins.
Superoxide dismutase 2 (<i>SOD2</i>)	NM_000636	Member of the iron/manganese superoxide dismutase family and associated with oxidative stress. It encodes a mitochondrial protein that binds to the superoxide byproducts of oxidative phosphorylation and converts them to hydrogen peroxide and diatomic oxygen.
Mitogen-activated protein kinase 14 (<i>MAPK14</i>)	NM_001315	Codes for a MAP-family member kinase that is activated by various environmental stressors and proinflammatory cytokines. It acts as an integration point for multiple biochemical signals, and is involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development. The substrates of this kinase include transcription regulator ATF2, MEF2C, and MAX, cell cycle regulator CDC25B, and tumor suppressor p53, which suggest the roles of this kinase in stress related transcription and cell cycle regulation, as well as in genotoxic stress response.
Chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha) (<i>CXCL1</i>)	NM_001511	Regulate cell trafficking of various types of leukocytes through interactions with a subset of 7-transmembrane, G protein-coupled receptors. Chemokines also play fundamental roles in the development, homeostasis, and function of the immune system, and they have effects on cells of the central nervous system as well as on endothelial cells involved in angiogenesis or angiostasis.
PTEN-induced putative kinase-1 (<i>PINK1</i>)	NM_032409	Gene Ontology: ATP binding, magnesium ion binding, nucleotide binding, protein serine/threonine kinase activity, transferase activity, protein kinase cascade, response to stress.
Diacylglycerol O-acyltransferase homolog 2 (mouse) (<i>DGAT2</i>)	NM_032564	DGAT is responsible for the synthesis of triglycerides. It catalyzes a reaction in which diacylglycerol is covalently joined to long chain fatty acyl-CoAs.
Tigger transposable element derived 3 (<i>TIGD3</i>)	NM_145719	The protein encoded by this gene belongs to the tigger subfamily of the pogo superfamily of DNA-mediated transposons in humans. These proteins are related to DNA transposons found in fungi and nematodes, and more distantly to the Tc1 and mariner transposases. They are also very similar to the major mammalian centromere protein B. The exact function of this gene is not known.

¹ See NCBI at <http://www.ncbi.nlm.nih.gov/Gene>

Materials and methods

Study population

The study population consisted of 398 apparently healthy subjects from eight different areas of inhabitation in Flanders, Belgium (Table 2), all recruited by the division of Bio-monitoring and Surveillance of the Flanders Environment and Health Study FLEHS. Study participants were aged 50 to 65 years at the time of inclusion. Prior to blood collection, informed consent was obtained from all individuals. Participants completed questionnaires on age, sex and smoking habits and donated single blood and urine samples for the measurement of health, exposure and early biological effect markers. All individuals were sampled once. The regions of inhabitation are characterized in Table 2 with respect to their environmental exposure features.

Figure 1. Map of Flanders representing the different regions of inhabitation of the study participants. The Antwerp and Gent harbors together represent the region 'Harbor'.

Table 2. Description of the represented regions in Flanders

Region	N _{inhabitants}	N _{subjects}	Description; most probable/abundant carcinogenic exposures
Antwerp	404.241	50	City and suburbs with exclusion of harbor area; PM, PAH
Harbor	130.064	41	Harbor areas of Antwerp and Gent combined (petrochemical and steel industries)
Fruit	95.829	35	Eight municipalities with >10 hectares/km ² apple and pear cultivation; pesticides, insecticides
Olen	68.068	39	Influenced by the presence of non-ferro industrial company Umicore; Pb, Cd, Zn
Gent	213.025	72	City and suburbs with exclusion of harbor area; PM, PAH
Incinerators	56.405	50	Surroundings of 11 incinerators; municipalities with their center <6 km from an incinerator (or <12 km in north-east direction); dioxins, hydrocarbons, (heavy) metals
Rural ¹	153.770	76	Twenty-four municipalities with inhabitant density <250/km ²
Albert canal	64.763	35	Presence of 6 chemical companies (BP-Chembel, Exxon Mobil, Dow Chemical, Borealis, T.C. Ham, Tessenderlo Chemie-LVM); benzene, Hg, Cu and other (heavy) metals

¹Municipalities with no registered emission source, >5% industry and no highway in the territory.

Gene expression analysis

For the gene expression analyses, blood samples of 2.5 ml were collected of each subject into PAXgene Blood RNA vacutainer tubes (PreAnalytix, Qiagen, Hilden, Germany) as this system accounts for immediate *ex vivo* preservation of the whole blood RNA. Total RNA was isolated and purified using the PAXgene Blood RNA kit according to the manufacturer's instructions. The use of this system provides high quality and highly pure total RNA and is of special interest for blood collection in human populations as it stabilizes and conserved the total RNA directly *ex vivo*.

cDNA was synthesized from 2 µg of total RNA according to the manufacturer's instructions using the BioRad iScript cDNA synthesis kit (Bio-Rad Laboratories, Hercules, CA). Aliquots were used for real-time PCR on the BioRad MyiQ iCycler Single Color real-time PCR detection system using iQTM SYBR[®] Green Supermix, containing iTaq Polymerase, dNTPs, SYBR Green I and buffers. Reactions were initiated for 3 min at 95°C to activate Taq polymerase, followed by 40 cycles of 15 seconds at 95°C and 45 seconds at 60°C. After each run, melting curve analysis was performed starting at 60°C with stepwise temperature elevations of 0.5°C every 10 seconds to check for nonspecific products. Beta-actin (*ACTB*) and cyclophilin A (*PPIA*) were included as reference genes (internal controls) as these genes have shown to perform best (most stably expressed and most resemblance to microarray derived results) in our previous analyses (data not reported). All reactions were performed in duplicate. In each run, negative controls (no template) and positive controls (a dilution series of a pooled sample, consisting of cDNA reverse transcribed from total RNA of 20 randomly selected subjects) were included to estimate PCR efficiency.

Primer sequences were as follows:

CYP1B1

5'-AGTGCAGGCAGAATTGGATCA-3' (forward) and 5'-GCGCATGGCTTCATAAAGGA-3' (reverse)

ATF4

5'-CTCCAGCGACAAGGCTAAGG-3' (forward) and 5'-GTTGTTGGAGGGACTGACCAA-3' (reverse)

MAPK14

5'-TGAAGACTGTGAGCTGAAGATTCTG-3' and 5'-CCACGTAGCCTGTTCATTTTCATC-3' (reverse)

SOD2

5'-ATCAGGATCCACTGCAAGGAA-3' (forward) and 5'-CGTGCTCCCACACATCAATC-3' (reverse)

CXCL1

5'-CCACTGCGCCCAAACC-3' (forward) and 5'-GCAGGATTGAGGCAAGCTTT-3' (reverse)

TIGD3

5'-GTGCTGGAACCTCTGGATGAG-3' (forward) and 5'-TTGCAGATGCGCGAGATCT-3' (reverse)

DGAT2

5'-GCACAGAGGCCACAGAAGTG-3' (forward) and 5'-CCCTCAACACAGGCATTTCG-3' (reverse)

PINK1

5'-AGCAGTCACTTACAGAAATCCAAGA-3' (forward) and 5'-GGTGAAGGCGCGGAGAA-3' (reverse)

ACTB

5'-CCTGGCACCCAGCACAAT-3' (forward) and 5'-GCCGATCCACACGGAGTACT-3' (reverse)

PPIA

5'-TTCCTGCTTTCACAGAATTATTCC-3' (forward) and 5'-GCCACCAGTGCCATTATGG-3' (reverse)

Ct values per subject and per gene were normalized by subtraction of the mean Ct value of *ACTB* and *PPIA*. Subsequently, $\Delta\Delta\text{Ct}$ values were calculated per subject and per gene relative to the normalized pooled reference sample within each real-time PCR run. This equals the following formula:

$$\Delta\Delta\text{Ct} = \text{pooled reference (Ct}_{\text{gene}} - \text{Ct}_{\text{mean ACTB-PPIA}}) - \text{individual (Ct}_{\text{gene}} - \text{Ct}_{\text{mean ACTB-PPIA}})$$

This generated one expression value per gene per individual study participant. All gene expression data are reported as $\Delta\Delta\text{Ct}$ values on log (base = 2) scale [19] and presented as average group values.

Measurement of early biological effect markers

Loss of DNA integrity was evaluated by the alkaline COMET assay as described by Singh *et al.* [20]. The slides were analyzed using an image analysis system of Metasystems (Altlussheim, Germany), viewing 50 cells per individual. Median percentages of DNA migration in the tail areas were determined and used as a measure of DNA damage. As positive control one slide with nuclei from deep frozen whole blood was added to each electrophoresis chamber. DNA migration in those positive controls had to be > 30%.

To investigate the frequencies of micronuclei (MN), the cytokinesis-block micronucleus assay was performed on whole blood cultures using standard procedures according to Fenech [21]. For each individual, 1000-2200 cells were evaluated for the presence of micronuclei using the Metafer automatic program from Metasystems (Altlussheim, Germany).

8-OH-dG in urine was measured by means of ELISA using the competitive immunosorbent assay (Gentaur, Brussels, Belgium), according to the manufacturer's instructions.

The serum protein level of the tumor marker p53 was analyzed using the enzyme immuno-metric assay (EIA) and Titerzyme EIA p53 (Assay Designs, Ann Arbor, MI, USA). CEA and PSA levels were measured using solid-phase chemiluminescent immunometric assay and Immulite 2000 CEA and PSA (DPC, Los Angeles, CA, USA). Samples were analyzed for PSA within 24 hours after collection.

Data analysis

Statistical testing was performed using SPSS 11.5 (SPSS Inc. Chicago, USA). Significance of differences in gene expression between groups of individuals inhabiting different regions and differences between groups of individuals with different smoking status (current versus former, current versus never and former versus never smokers) and other variables, the exposure and effect biomarkers, was tested by means of one-way analysis of variance with *post hoc* Bonferroni correction. For examination of inter-variable correlations at the level of all individuals (no division based on region of inhabitation or smoking status), Pearson correlation analysis was applied. P-values <0.05 are considered statistically significant.

Results

The effects of environmental exposures on the expression of eight key genes were analyzed by means of real-time PCR in blood cells from human subjects inhabiting eight different populations across Flanders, each with presence of select environmental carcinogenic pollutants (Table 2). *CYP1B1*, *ATF4*, *SOD2*, *MAPK14*, *CXCL1*, *DGAT2*, *TIGD3* and *PINK1* expressions were analyzed in 398 individuals. As a measure for the total gene expression profile, the mean and sum of the eight gene expressions was calculated per individual.

Table 3 depicts the characteristics of the study population according to the region of inhabitation and of the population as a whole. Among all 398 study participants, 68 (17%) were current tobacco smokers, 107 (27%) were former smokers, 184 (46%) had never smoked tobacco and for 39 (10%) the smoking status was unknown. Individuals with unknown smoking status were excluded from the analysis.

Study participants according to region did not differ significantly with respect to age, distribution of males and females and smoking status. *A priori* tobacco smoking was suspected to be an important confounding factor for the analysis of the region-based effects in relation to the burden of environmental carcinogen exposure. Therefore, smoking-dependent effects were analyzed prior to the region-dependent effects. The majority of individuals were classified as never smokers according to self-reported smoking status. The former smokers self-reported regular smoking in the past (no information was available concerning the time of smoking cessation), but to have quit at the time of the research. At this moment, no data on cotinine levels were available to confirm smoking status.

Table 3. Characteristics of the study population according to region of inhabitation

	Antwerp	Harbor	Fruit	Olen	Gent	Incinerator	Rural	Canal	Total
N of subjects	50	41	35	39	72	50	76	35	398
Age	60.2 ± 0.6	58.4 ± 0.7	58.6 ± 0.9	58.9 ± 0.7	58.9 ± 0.6	59.0 ± 0.6	59.3 ± 0.5	60.1 ± 0.6	59.2 ± 0.2
Gender									
Males	30 (60%)	20 (51%)	19 (54%)	17 (44%)	38 (53%)	26 (52%)	43 (57%)	19 (54%)	212 (53%)
Females	20 (40%)	21 (49%)	16 (46%)	22 (56%)	34 (47%)	24 (48%)	33 (43%)	16 (46%)	186 (47%)
Smoking status									
Current	15 (30%)	5 (12%)	2 (6%)	6 (15%)	11 (15%)	11 (22%)	12 (16%)	6 (17%)	68 (17%)
Former	14 (28%)	10 (24%)	9 (26%)	14 (36%)	19 (26%)	16 (32%)	17 (22%)	8 (23%)	107 (27%)
Never	20 (40%)	18 (44%)	13 (37%)	17 (44%)	35 (49%)	23 (46%)	41 (54%)	17 (49%)	184 (46%)
Unknown	1 (2%)	8 (20%)	11 (31%)	2 (5%)	7 (10%)	0 (0%)	6 (8%)	4 (11%)	39 (10%)
Cells with MN	8.56 ± 0.74	5.42 ± 1.14	14.5 ²	6.69 ± 0.60	7.08 ± 0.47	9.23 ± 0.96	7.78 ± 0.77	7.67 ± 1.10	7.76 ± 0.30
MN %	9.51 ± 0.84	5.67 ± 1.17	15.4 ²	7.24 ± 0.68	7.80 ± 0.53	10.36 ± 1.09	8.66 ± 0.89	8.94 ± 1.60	8.63 ± 0.36
8-OH-dG¹	15.04 ± 0.73	14.08 ± 0.80	16.91 ± 1.27	15.00 ± 1.14	15.64 ± 0.68	18.65 ± 1.22	14.90 ± 0.67	15.66 ± 0.76	15.72 ± 0.32
COMET(count)	209.5 ± 5.4	202.0 ± 6.1	225.0 ²	200.3 ± 7.6	208.5 ± 5.5	211.5 ± 3.9	212.0 ± 5.7	220.0 ± 1.3	209.6 ± 2.2
COMET (P90)	8.96 ± 0.43	12.90 ± 1.23	8.60 ²	8.79 ± 0.40	8.40 ± 0.18	8.04 ± 0.23	8.60 ± 0.22	8.55 ± 0.29	8.72 ± 0.13
COMET(median)	1.77 ± 0.22	3.91 ± 0.63	1.17 ²	1.61 ± 0.19	2.30 ± 0.15	1.76 ± 0.13	2.29 ± 0.12	1.94 ± 0.17	2.07 ± 0.07
PSA (ng/ml)	1.51 ± 0.19	1.29 ± 0.29	1.16 ± 0.31	1.13 ± 0.23	2.21 ± 0.79	0.92 ± 0.17	1.27 ± 0.16	1.43 ± 0.32	1.42 ± 0.16
CEA (ng/ml)	2.69 ± 0.51	1.96 ± 0.38	0.98 ²	2.29 ± 0.33	2.52 ± 0.28	2.66 ± 0.33	2.59 ± 0.34	1.62 ± 0.16	2.43 ± 0.14
PS3 (pg/ml)	44.4 ± 17.2	218.3 ± 131.9	68.7 ± 35.4	74.9 ± 37.8	94.0 ± 35.1	74.1 ± 37.2	103.3 ± 44.1	42.5 ± 19.8	80.5 ± 14.0

Excluded are the individuals of which no smoking status was known or reported as smoking is expected a priori to have a confounding effect.

Presented are mean values ± standard errors of the mean.

¹ Values in µg/g creatinine.

² Value based on one individual.

Statistical analysis of gene expressions between current, former and never smokers revealed *CYP1B1* and *ATF4* expression levels to be significantly different (analyzed by t-test, Figure 2). For *PINK1* the difference between current and never smokers was nearly significant ($p=0.077$). Although not significant, *MAPK14* also showed an increased expression in current smokers compared to former smokers, and former smokers compared to never smokers. Thus, for several genes their expression differs between smokers and non-smokers (*i.e.* never smokers and former smokers). Since the effects on gene expression caused by smoking may obscure the effects due to the region of inhabitation, expressions were investigated in both non-smokers only and in all study subjects (excluding the individuals with unknown smoking status).

Figure 2. Expressions of genes in current, former and never smokers.* $p=0.02$ for the difference between current and never smokers.** $p=0.000$ for the difference between current and never smokers and $p=0.005$ for the difference between current and former smokers.

Figure 3 depicts the map of Flanders with bar charts of the average gene expressions per region. Although smoking was considered a confounding factor for gene expression, observed differences in gene expression levels appeared independent from effects of smoking behavior on these gene expressions. The regions that show the most distinct gene expression profiles as compared to the total population average are Olen (well above the population average) and Fruit and Gent (well below the population average). In a one-way ANOVA analysis with a *post hoc* Bonferroni test, all genes appeared to significantly differ in expression between two or more regions ($p<0.001$), except for *CXCL1* ($p=0.06$) and *DGAT2* ($p=0.14$). Based on the individual gene expressions as well as the mean/sum of all gene expressions, the region that most significantly differed from any of the other regions was Olen, showing the most significant differences with the fruit cultivation area (Fruit) and Gent ($p<0.001$).

Figure 3. Gene expressions per region presented as $\Delta\Delta C_t$ values, relative to the total population averages. Error bars represent standard errors of the mean. Non-smokers include never smokers and former smokers

Ranking the regions on the basis of the highest overall effect, *i.e.* the highest sum of gene expressions, the following order was observed (from high to low):

Olen>Incinerators> Antwerp>Canal>Rural>Harbor>Gent>Fruit

Due to logistic and organizational issues, samples were collected at two or three dates in each region, between September 29th 2004 and June 6th 2005. Therefore, we next investigated whether day or season of blood collection had a confounding effect on the gene expressions observed in the different regions (Figure 4). As analyzed by one-way ANOVA with *post hoc* Bonferroni test, the gene expressions appeared to be influenced by the day or season of blood collection, showing (statistically significant) differences between two or more blood collection days within one region (Figure 4; only for non-smokers). In this figure, the results are arranged on the season of blood collection: fall (Sep 27-Dec 21), winter (Dec 21-March 21) or spring (March 21-June 6). No sampling occurred during summer months. Significant differences between seasons were dependent on the region and the gene measured, *e.g.* in the Harbor region the expression of four genes (*CYP1B1*, *ATF4*, *MAPK14* and *PINK1*) was influenced by the season, whereas in the Rural region no significant seasonal effects were observed. Although within the regions season differences are observed, overall no consistent seasonal effect was observed and season differences were not always significant.

Figure 4. Gene expressions of all non-smokers (i.e. never and former smokers) per region of inhabitation, according to season of blood collection. Within a region, the data points are in chronological time order. Expression values are given as $\Delta\Delta C_t$ values relative to the average of the total population. Error bars represent standard errors of the mean. A=Antwerp, H=Harbor, F=Fruit, O=Olen, G=Gent, I=Incinerators, R=Rural, C=Canal.

Hierarchical clustering maps of all genes and of genes with other parameters are shown in Figure 5. The analyses grossly revealed a distinction between two clusters of regions; one including Antwerp, Canal, Incinerators and Rural, and the other including Fruit, Gent and Harbor. Olen and Fruit and Gent were again the most deviating regions, based on these cluster maps.

Figure 5. Hierarchical clustering (average standardized Euclidean) of gene expressions according to regions (A) and of genes with other parameters (the effect markers, tumor markers) (B). Values are from non-smokers and relative to the average of the total population. COM-COUNT = COMET count, COMP90 = COMET P90, COMMED = COMET median.

Pearson correlation analyses were performed to investigate associations between individual gene expressions, *i.e.* gene expression values per study participant regardless of region of inhabitation, and the data on DNA damage (COMET), micronuclei and the three tumor markers PSA, CEA and p53. All significant correlations for are presented in Figure 6. Correlation analyses were performed both including and excluding the current smokers. Both analyses returned significant correlation of the majority of genes with COMET (count and P90) and p53 protein level. The associations as presented in Figure 6 were negative for the correlation of gene expression with COMET count and positive for COMET P90. Correlations of any of the gene expressions with any of the tumor markers were negative, except for correlation of *CYP1B1* gene expression with serum CEA. No associations were found between any of the genes and the number of cells with MN or the MN frequencies.

Figure 6. Significant correlations of gene expressions and other parameters ($p < 0.05$) excluding (A) and including current smokers (B). Bold lines represent correlations with a correlation coefficient ≥ 0.15 .

Discussion

The Flanders Environment and Health Study (FLEHS) was initiated by the Flemish government in 2001, with the objective to evaluate the relation of public health and environmental exposure and with the aim to imply the study results for environmental risk assessment. Flanders is a typical region with different types of environmental situations such as urban areas, dense traffic, intensive agriculture and industry. In the present study, we investigated the expression of eight key genes across the general adult and apparently healthy Flanders population. Real-time PCR was used to study gene expression in peripheral blood cells of these environmentally-exposed humans. The eight genes that were monitored have been identified as promising biomarkers for environ-

mental carcinogenesis in two previous population studies: one that investigated differential gene expression in cigarette smokers versus non-smokers (**Chapter 3**) and another study among children in the Czech Republic exposed to elevated concentrations of air pollution as compared to children from a relatively clean area (**Chapter 4**). Both studies delivered sets of genes, verified by real-time PCR, of which the expression differed significantly between the exposed and non-exposed individuals and also genes that correlated with established biomarkers of early biological effect (micronuclei). From the most relevant results of both studies, the presently monitored eight genes were selected. Next to gene expression for the eight individual genes, an expression value was generated to represent the total gene expression profile by calculation of the sum of the eight individual genes. This was done as the combination of genes might be a more relevant or stable measure, considering inter-individual variability in the expression of the single genes by e.g. polymorphisms.

Data analysis was commenced with investigating an *a priori* hypothesized important confounder; tobacco smoking. Therefore, first the effects of smoking on the gene expressions were analyzed. Statistical analyses of the effects of tobacco smoking on the gene expression revealed *CYP1B1* gene expression to significantly differ between current, former and never smokers and *ATF4* expression between current and never smokers. For *CYP1B1* gene expression in peripheral blood cells, an association with smoking has been observed previously [17] and has been found to significantly positively correlate with plasma cotinine levels in smokers [22]. Elevated *CYP1B1* gene expression due to smoking, PAH or dioxin exposure has also been described in human blood, lung and liver cells [23-25]. *ATF4* activation by tobacco smoke exposure has not been reported up to now, but has been found in HepG2 cells upon arsenite exposure [26]. Since only two out of eight genes significantly differed between smokers and non-smokers, the influence of tobacco smoking on the analysis of the region-based effects appears limited.

Although smoking was proven a confounding factor for gene expression of *CYP1B1* and *ATF4*, observed differences in gene expression levels between the different regions appeared independent from effects of smoking behavior on these gene expressions (Figure 3). One-way ANOVA analysis of all non-smokers with a *post hoc* Bonferroni test, revealed all of the genes to significantly differ between two or more regions ($p < 0.001$), except for *CXCL1* ($p = 0.06$) and *DGAT2* ($p = 0.14$). As the scheme below shows, the ranking order of the regions with regard to the magnitude of the gene expression also differed per gene in an analysis of non-smokers only. Therefore it may be relevant to regard the gene expression profile as a whole (the sum or mean of the eight gene expressions) as this reflects the average or overall effect.

<i>CYP1B1</i>	Olen>Antwerp>Incinerators>Canal>Rural>Harbor>Gent>Fruit
<i>ATF4</i>	Antwerp>Olen>Canal>Incinerators>Rural>Harbor>Fruit>Gent
<i>MAPK14</i>	Olen>Incinerators>Canal>Antwerp>Harbor>Rural>Fruit>Gent
<i>SOD2</i>	Olen>Incinerators>Canal>Harbor>Antwerp>Rural>Fruit>Gent
<i>CXCL1</i>	Antwerp>Olen>Rural>Harbor>Incinerators>Gent>Fruit>Canal
<i>DGAT2</i>	Incinerators>Olen>Antwerp>Harbor>Gent>Fruit>Canal>Rural
<i>TIGD3</i>	Olen>Canal>Antwerp>Rural>Harbor>Gent>Incinerators>Fruit
<i>PINK1</i>	Incinerators>Rural>Olen>Canal>Antwerp>Gent>Fruit>Harbor
SUM/MEAN	Olen>Incinerators>Antwerp>Canal>Rural>Harbor>Gent>Fruit

Based on the individual gene expressions as well as the mean/sum of all gene expressions, the most deviating region is Olen, as it returned the most significant differences with any of the other regions, in particular the regions Harbor, Fruit, Gent and Rural (see also Figure 3). In line with this observation, the hierarchical clustering analyses returned two gross clusters of regions, based on gene expressions only (Figure 5A) as well as on all parameters (including biological effect markers and tumor markers; Figure 5B). Regions Olen, Antwerp, Incinerators and Canal were one cluster with an above population average expression, whereas Fruit, Harbor, Gent and Rural shaped the other cluster with a below population average expression. The regions were selected based on environmental monitoring data and known local presence of chemicals and carcinogens in the environment, such as particles and PAH from traffic or combustion in cities and incinerator areas, pesticides and insecticides in the fruit cultivation region and heavy metals like cadmium and lead in the surroundings of Olen. However, no individual measurements of biomarkers of exposure or internal dose were available at the time of analysis of the present data. In the (near) future, additional information on internal concentrations of PAHs, PCBs, dioxin and heavy metals will become available for the individuals currently analyzed. This information will provide the possibility for a more focused analysis of the continuum of biomarkers from exposure up to tumor markers. The ability to quantify internal exposure in the individuals will provide a better fundament for the analysis of the environmentally induced effects than the current approach at group level, namely based on the region of inhabitation. A possible *a priori* assumption was that regions with theoretically similar exposures, such as Harbor and Canal, Antwerp and Gent and Fruit and Rural (Table 2), would display similar effects on gene expression and other biomarkers. This is however not observed from the clustering in Figure 5 and may be due to differences in individual exposures in humans inhabiting the same region. This finding illustrates the importance of monitoring internal exposure levels of environmental carcinogens, to indicate potential differences in individual exposures and internal doses. This enables a more directed analysis of high versus low environmental exposure as measurements of multiple exposure markers will provide a better estimation of the effects of the complex mixtures of chemicals that might occur in the different regions.

Besides tobacco smoking, another factor of influence on gene expression was identified in the day or season of blood collection. Figure 4 shows the expressions of the genes according to region of inhabitation, where divisions within regions were made based on the season on blood collection. In each region, sampling occurred at two or three different dates in fall, winter and/or spring. In a number of cases (gene expressions in a certain region), significant differences were observed between the two or three dates (Figure 4). However, expression of the genes according to region did not show to be influenced systematically by the season. Since sampling in the different regions occurred throughout different seasons and not all inter-region differences between seasons were significant, we considered correction of the dataset for the effects of the sampling moment to complicate analysis of the goal; differences between regions. Thus, although the season of sampling appears to have an effect, from the present results it does not seem to have major impact. However, an in depth analysis of this issue will be subject to future research. In conclusion, it appears feasible to significantly separate different regions of inhabitation with different characteristics of environmental exposure, using gene expression or a gene expression profile. However, additional information on biomarkers of

internal dose, reflecting individual exposure, is needed to fully interpret the data and to further strengthen the conclusion.

Besides analysis of the regional effects, correlations of gene expressions with the other parameters were investigated across the total population at individual level and regardless of the region of inhabitation (Figure 6). These analyses were performed both including and excluding the current smokers. Some differences between these two can be observed, which suggest tobacco smoking to exert an influence on the correlations. Significant associations of individual genes with more than one biomarker along the biomarker continuum, suggests gene expression to be a rather comprehensive event. Five out of eight gene expressions (*MAPK14*, *SOD2*, *CXCL1*, *TIGD3* and *DGAT2*) as well as the sum (or mean) of all gene expressions were found to correlate significantly and negatively with p53 serum protein levels ($p < 0.05$, Figure 5). This was observed both excluding and including the current smokers, however the correlations were stronger for the non-smokers only. This is due to the fact that serum p53 levels in current smokers are significantly lower than in former and never smokers (data not reported). A similar observation has been done by Rössner *et al.*, who found that personal exposures to occupational PAHs in coke-oven workers negatively correlated with levels of serum p53 protein [27]. Also, Stierum *et al.* observed elevated p53 expression after treatment of isolated peripheral blood cells with low(er) doses of *anti*-BPDE [28]. Concluding, gene expression appears to correlate with biomarkers of exposure, of early biological effect and with tumor markers in apparently healthy humans. Different genes correlated with different (combinations of) biomarkers.

For comparison of the classical markers with the gene expressions with respect to sensitivity, it was observed that the majority of gene expressions significantly differed between two or more regions, whereas the micronuclei frequencies and total COMET count did not. COMET P90 and median values and 8-OH-dG did differ significantly between two or more regions, however not as significant as the gene expressions did. In addition, the magnitudes of the differences between regions displaying the lowest and the highest values for the gene expressions and the other parameters were observed to be quite similar; 1.2 (*DGAT2*) - 2.0 (*ATF4*) against 1,10 (COMET count) – 2.43 (COMET median) respectively.

Regarding specificity, tobacco smoking was observed to be of limited influence on the gene expressions as only two out of eight genes of the profile investigated, significantly differed between smokers and non-smokers. Significant differences were observed between smokers and non-smokers for micronuclei frequency, CEA and p53, which therefore may appear of lower specificity for environmental exposure than gene expression. Therefore, our results suggest the gene expression (profile) to be no less sensitive or specific than the classical biomarkers measured.

Gene expressions have been monitored in peripheral blood from human populations before, however to our knowledge they have not been associated with classical biomarkers of effect. Wu *et al.* have described gene expression profiling in a Taiwanese population exposed to arsenic-pollution through drinking water [29]. They found differential expression between groups with low, intermediate and high arsenic exposure, in particular inflammation-related genes showing up-regulation with increasing exposure. In addition, gene expression profiling has been used to examine the risks of occupational exposures to benzene [30] and metal fumes [31]. However these studies did not com-

prise as many study subjects, as the present study. Therefore, the current study is the first to investigate differential gene expression in a large population.

The results of this study revealed great differences in gene expression between humans living in different regions in Flanders, especially between Olen and the fruit cultivation area. This suggests the different groups to suffer different health risks. As all monitored gene expressions (are expected to) respond to increasing exposures with an up-regulation, this suggests the humans inhabiting the region of Olen to be at the highest and the human from the fruit cultivation region to be at the lowest health risk of all populations in this study.

In conclusion, overall, this study shows that gene expressions or a gene expression profile can significantly segregate human populations according to the region of inhabitation, revealing the greatest difference between the regions of Olen and those of fruit cultivation and Gent. In addition, gene expression was found to significantly correlate with DNA strand breaks (COMET) and with tumor markers. Gene expression profiling therefore appears promising for application to the analysis of the risks of environmental exposures in humans.

Acknowledgements

The authors thank the volunteers for participating and the nurses for sample collection. This study was conducted as part of and financially supported by the Flanders Environment and Health Study.

References

1. Boyle, P. and Ferlay, J. (2005). Cancer incidence and mortality in Europe, 2004. *Ann Oncol* 16, 481-488.
2. Czene, K., Lichtenstein, P. and Hemminki, K. (2002). Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer* 99, 260-266.
3. Le Marchand, L. (2005). The predominance of the environment over genes in cancer causation: implications for genetic epidemiology. *Cancer Epidemiol Biomarkers Prev* 14, 1037-1039.
4. Luch, A. (2005). Nature and nurture - lessons from chemical carcinogenesis. *Nat Rev Cancer* 5, 113-125.
5. Higginson, J. (1993). Environmental carcinogenesis. *Cancer* 72, 971-977.
6. Weinstein, I.B., Santanella, R.M. and Perera, F.P. The molecular biology and molecular epidemiology of cancer., in: Greenwald, P. (Ed.), *The Science and Practice of Cancer Prevention and Control*, Marcel-Dekker, New York, NY, 1993.
7. Bonassi, S., Ugolini, D., Kirsch-Volders, M., Stromberg, U., Vermeulen, R. and Tucker, J.D. (2005). Human population studies with cytogenetic biomarkers: review of the literature and future perspectives. *Environ Mol Mutagen* 45, 258-270.
8. Orphanides, G. (2003). Toxicogenomics: challenges and opportunities. *Toxicol Lett* 140-141, 145-148.
9. Simmons, P.T. and Portier, C.J. (2002). Toxicogenomics: the new frontier in risk analysis. *Carcinogenesis* 23, 903-905.
10. Aardema, M.J. and MacGregor, J.T. (2002). Toxicology and genetic toxicology in the new era of "toxicogenomics": impact of "-omics" technologies. *Mutat Res* 499, 13-25.
11. Afshari, C.A., Nuwaysir, E.F. and Barrett, J.C. (1999). Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. *Cancer Res* 59, 4759-4760.
12. Gerhold, D., Lu, M., Xu, J., Austin, C., Caskey, C.T. and Rushmore, T. (2001). Monitoring expression of genes involved in drug metabolism and toxicology using DNA microarrays. *Physiol Genomics* 5, 161-170.
13. Hamadeh, H.K., Bushel, P., Paules, R. and Afshari, C.A. (2001). Discovery in toxicology: mediation by gene expression array technology. *J Biochem Mol Toxicol* 15, 231-242.
14. Lobenhofer, E.K., Bushel, P.R., Afshari, C.A. and Hamadeh, H.K. (2001). Progress in the application of DNA microarrays. *Environ Health Perspect* 109, 881-891.
15. Lockhart, D.J. and Winzler, E.A. (2000). Genomics, gene expression and DNA arrays. *Nature* 405, 827-836.
16. van Leeuwen, D.M., Gottschalk, R.W., van Herwijnen, M.H., Moonen, E.J., Kleinjans, J.C. and van Delft, J.H. (2005). Differential gene expression in human peripheral blood mononuclear cells induced by cigarette smoke and its constituents. *Toxicol Sci* 86, 200-210.
17. van Leeuwen, D.M., van Agen, B., Gottschalk, R.W.H., Vlietinck, R., Gielen, M., van Herwijnen, M.H.M., Maas, L., Kleinjans, J.C.S. and van Delft, J.H.M. (2005). Cigarette smoke-induced differential gene expression in blood cells from monozygotic twin pairs (in press).
18. van Leeuwen, D.M., van Herwijnen, M., Pedersen, M., Knudsen, L., Kirsch-Volders, M., Sram, R., Staal, Y.C.M., Bajak, E., van Delft, J.H.M. and Kleinjans, J.C.S. (2005). Genome-wide differential gene expression in children exposed to air pollution in the Czech Republic. *Mutat Res* 600, 12-22.
19. Livak, K.J. and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods* 25, 402-408.
20. Singh, N.P., McCoy, M.T., Tice, R.R. and Schneider, E.L. (1988). A simple technique for quantitation of low levels of DNA damage in individual cells. *Exp Cell Res* 175, 184-191.
21. Fenech, M. (2000). The in vitro micronucleus technique. *Mutat Res* 455, 81-95.
22. Lampe, J.W., Stepaniants, S.B., Mao, M., Radich, J.P., Dai, H., Linsley, P.S., Friend, S.H. and Potter, J.D. (2004). Signatures of environmental exposures using peripheral leukocyte gene expression: tobacco smoke. *Cancer Epidemiol. Biomarkers Prev* 13, 445-453.
23. Chang, T.K., Chen, J., Pillay, V., Ho, J.Y. and Bandiera, S.M. (2003). Real-time polymerase chain reaction analysis of CYP1B1 gene expression in human liver. *Toxicol Sci* 71, 11-19.
24. Han, W., Pentecost, B.T., Pietropaolo, R.L., Fasco, M.J. and Spivack, S.D. (2005). Estrogen receptor alpha increases basal and cigarette smoke extract-induced expression of CYP1A1 and CYP1B1, but not GSTP1, in normal human bronchial epithelial cells. *Mol Carcinog* 44, 202-211.
25. Hu, S.W., Chen, C.C., Kuo, C.Y., Lin, W.H. and Lin, P. (2006). Increased cytochrome P4501B1 gene expression in peripheral leukocytes of municipal waste incinerator workers. *Toxicol Lett* 160, 112-120.
26. Ord, D. and Ord, T. (2005). Characterization of human NIPK (TRB3, SKIP3) gene activation in stressful conditions. *Biochem Biophys Res Commun* 330, 210-218.
27. Rossner, P., Jr., Binkova, B. and Sram, R.J. (2003). The influence of occupational exposure to PAHs on the blood plasma levels of p53 and p21WAF1 proteins. *Mutat Res* 535, 87-94.

28. Stierum, R.H., van Herwijnen, M.H., Pasman, P.C., Hageman, G.J., Kleinjans, J.C. and van Aken, B. (1995). Inhibition of poly(ADP-ribose) polymerase increases (+/-)-anti-benzo [a]pyrene diol-epoxide-induced micronuclei formation and p53 accumulation in isolated human peripheral blood lymphocytes. *Carcinogenesis* 16, 2765-2771.
29. 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.
30. Forrest, M.S., Lan, Q., Hubbard, A.E., Zhang, L., Vermeulen, R., Zhao, X., Li, G., Wu, Y.Y., Shen, M., Yin, S., Chanock, S.J., Rothman, N. and Smith, M.T. (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.
31. Wang, Z., Neuburg, D., Li, C., Su, L., Kim, J.Y., Chen, J.C. and Christiani, D.C. (2005). Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. *Environ Health Perspect* 113, 233-241.

Chapter 7

Summary and General Discussion

The research described in this thesis was conducted with the primary objective to develop gene expression profiles as a new biomarker of early response to chemical carcinogen exposure for eventual application in a human biomonitoring study. The range of biomarkers that is currently applied in the field of molecular epidemiology does not yet comprise biomarkers that provide generic, biologically relevant and mechanistic information. As the response of a cell to xenobiotic exposure, such as to chemical carcinogens, is known to be partly the result of modulations of the expression of genes, monitoring those altered gene expression profiles is considered a promising approach to the identification of new biomarkers. With regard to the conceptual basis of biomarkers (Figure 2 in **Chapter 1**), gene expression is suggested to be an early biological response to xenobiotic exposure, in between biomarkers of exposure and those of early biological effect, as it is regarded to mark early effects that are in principle reversible. To evaluate the potential of gene expression profiles as a new biomarker, *in vitro* experimental studies as well as *in vivo* validation and case studies were performed, investigating mainly surrogate target tissue, but also target tissue for environmental exposures. Peripheral blood cells (PBC) were used throughout the major part of the research as this is the generally accepted surrogate target tissue in molecular epidemiology, representing an easily accessible biological matrix which can be obtained from humans with little invasiveness for the purpose of application in human biomonitoring studies. In these, often large scale studies in healthy human individuals, the target tissues such as airway epithelium cells, can not be obtained for ethical reasons.

The research was initiated by an exploratory *in vitro* study as described in **Chapter 2**, with the aim to examine modulations in gene expression in isolated (quiescent) peripheral blood mononuclear cells (PBMC). The transcriptomic response of PBMC to cigarette smoke condensate (CSC) or to its individual constituents benzo[a]pyrene (BaP), 4-aminobiphenyl (4-ABP), NNK (4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone) and hydrogen peroxide (H_2O_2) was investigated using microarray technology dedicated to application in toxicological studies. The exposures clearly induced significant modulations in gene expression, both up- and down-regulation. This study knew two objectives. First, a set of 16 genes was identified that responded to all agents, which therefore might be considered to represent a general response to toxicant exposure. However, when studying the expression of the individual genes, none of those appeared to show a consistent modulation in terms of the direction of regulation. Therefore, this set of genes was considered not useful as a generic biomarker. The second objective focused on identification of the most sensitive responses to each agent individually, defined by the genes showing (highly) significant differential regulation at the lowest or medium concentrations. This delivered different sets of genes for each agent, where CSC induced the most genes as compared to the individual carcinogens. This was in accordance with *a priori* expectations, since CSC comprises a large number of toxicants and carcinogens prone to affect gene expression. The expression of a selection of genes was verified by means of an independent assay, SYBR Green I real-time polymerase chain reaction (PCR), which confirmed the observations from the microarray analyses. Analysis of affected functionalities, namely overrepresented Gene Ontology terms for biological processes or molecular functions within the significantly responding gene lists, was performed using the online tool Expression Analysis Systematic Explorer (EASE). Immune response and defense response were found to be the biological processes most significantly overrepresented. In conclusion, this *in vitro* study demonstrated the feasibility of gene expres-

sion modulation by CSC and its constituents, resulting in lists of sensitively responding genes that may provide a basis for the selection of future candidate biomarker genes, which obviously needed to be verified by *in vivo* studies.

The *in vitro* study was succeeded by the first of two concise *in vivo* human biomonitoring studies which is the subject of **Chapter 3**. Peripheral blood from monozygotic twin pairs discordant for cigarette smoking, *i.e.* each twin pair consisted of a smoker and a non-smoker, were analyzed for effects on the transcriptome by smoking, again assessed by means of analysis on microarrays dedicated to toxicologically relevant genes. Identical twins were studied to reduce the possible influence from (differences in) genetic background on gene expression. The PAXgene Blood RNA system was applied for sample collection, as in this study humans were sampled at a location where no direct and standardized isolation of PBMC was possible. Storage and conservation of biological materials is a practical issue in human biomonitoring, especially in the current research which concentrated on RNA. The PAXgene system provided RNA stabilization and conservation; in this way, the gene expression profile in the blood was 'frozen' instantly *ex vivo*. The system delivered high quality RNA. This study was performed with two objectives, including an approach to benchmark or 'phenotypically anchor' the gene expressions by comparison with an established and validated biomarker of biological effective dose: aromatic DNA adducts. First, through the combination of results from several statistical tests, the most significantly differentially expressed genes in smokers as compared to non-smokers were identified as those genes with significant results in all of the tests; *ATF4*, *SOD2* and *MAPK14*. These differential gene expressions were confirmed by means of real-time PCR. In addition, real-time PCR was used to analyze expression of *CYP1B1* and *SERPINB2*, two genes that did not show significance from the microarray analyses. The results demonstrated significant up-regulation of the genes in smokers compared to non-smokers, which is in agreement with other studies (*CYP1B1*) and the previous study in cigarette smoke-exposed PBMC *in vitro* (*SERPINB2*, **Chapter 2**). The second approach was to investigate associations between individual gene expressions and DNA adduct levels. Correlation analysis returned seven gene expressions to correlate strongly and significantly with the DNA adduct levels. Again, SYBR Green I real-time PCR was used to confirm the microarray results. Real-time PCR did however not confirm the correlations of gene expressions with DNA adduct levels. Therefore, these genes were no longer considered.

The study as described in **Chapter 4** was directed toward analysis of the effects of cigarette smoke exposure at the target tissue level. The objective was to compare responses at the gene expression level in surrogate and target tissue by investigating differential gene expressions in primary normal human bronchial epithelial (NHBE) cells *in vitro* and comparing the findings with observations from the previous *in vitro* PBMC study (**Chapter 2**) and the smokers versus non-smokers *in vivo* study (**Chapter 3**). NHBE cells were exposed to different concentrations of CSC for various periods. Gene expression was analyzed using Agilent Human 22k oligonucleotide microarrays which represent more than 20.000 annotated gene sequences. A set of genes was identified that showed similarly significant responses to the exposures in both surrogate and target tissue *in vitro*. Also, some genes were observed to respond in target tissue as well as in blood from smokers *in vivo*. Cytochrome P450 1B1 (*CYP1B1*) was the only gene that showed significant expression responses in blood cells *in vitro* and *in vivo* as well as in NHBE cells *in vitro*. Besides differences in cell types being investigated, the use of dif-

ferent types/platforms of microarray might have contributed to this. Up-regulation of *CYP1B1* is in line with literature on the over-expression of this gene upon chemical or genotoxic exposure in both cell types. In conclusion, cigarette smoke induced effects in NHBE and these confirm to a certain extent the main effects found in peripheral blood cells previously.

A second population study, conducted in the Czech Republic, is described in **Chapter 5**. Two different features to this study, as compared to the previously described studies, are the investigation of children instead of adults and the chemical exposure being air pollution. To enrich the search for possible biomarker genes for eventual application in a more extensive biomonitoring study, we included this study as the exposure to air pollution represents a more general and abundant exposure to a complex mixture of chemical carcinogens than cigarette smoke. Children residing in the area of Teplice in the Czech Republic were investigated as this area is known for its (historically) severe high levels of air pollutants like (carcinogenic) polycyclic aromatic hydrocarbons (PAH). To analyze differential gene expression, Agilent Human 22k oligonucleotide microarrays were used. Genome-wide gene expression in blood of the Teplice children was studied compared to that in children inhabiting a relatively clean area in the Czech Republic, Prachatice. In addition, the gene expressions were investigated for association with individual micronuclei frequencies, which represent a biomarker of early biological effect. Both objectives resulted in clearly defined gene lists which were analyzed for overrepresentation of biological processes and molecular functions. Because in this study the exposure to pollutants is relatively low as compared to *e.g.* occupational carcinogen exposure or smoking, it was anticipated to find relatively small numbers of significantly modulated gene expressions. The analyses were therefore concentrated at pathway level; on finding overrepresented biological processes and molecular functions within the lists of significantly responding genes. Considerable numbers of genes were identified which expression differed significantly between both groups of children. The most significantly enriched biological processes within these lists were related to nucleosome and chromatin structure and maintenance (assembly/disassembly), which suggests deregulation of the primary structural unit of the DNA. Also, many gene expressions were observed to correlate significantly with micronuclei frequencies. Eventually, a set of five genes was selected from this study for real-time PCR confirmation: *CXCL1*, *DGAT2*, *TIGD3*, *PINK1* and *FXN*, all of which except *FXN* were confirmed. Criteria for selection were a statistically significant differential expression in Teplice children as compared to Prachatice children as well as a significant correlation of the expression with micronuclei frequencies. In conclusion, results of this study suggest that gene expression profiling appears a feasible and promising tool for application in the environmental health study setting. Therefore, it appears relevant to monitor more populations for transcriptomic responses to environmental carcinogenic contaminants.

Finally, **Chapter 6** describes the application of a gene expression profile, retrieved from the previously conducted studies, in human biomonitoring. A large-scale biomonitoring study was performed on differential gene expressions in peripheral blood from selected human adult populations in Flanders facing differential environmental exposures as a result of the area of inhabitation. Eight regions with characteristic theoretical environmental burdens were included; cities with dense traffic, surroundings of incinerators or industries as well as areas of fruit cultivation and rural regions. Because of the required numbers of study participants in this large-scale biomonitoring, this study

comprised nearly 400 individuals, real-time PCR instead of microarray technology was used for gene expression analysis. Throughout the research presented in this thesis, microarray analysis was performed for development of the gene expression profile (specific set of genes) as it is monitored in the study described in this chapter. Based on the previous studies in human populations exposed to cigarette smoke and air pollution, as described in **Chapters 3** and **5** respectively, eight genes were selected to be monitored in the Flemish population: *CYP1B1*, *ATF4*, *MAPK14* and *SOD2* based on the results of the study on smoking-discordant twins and *CXCL1*, *TIDG3*, *DGAT2* and *PINK1* from the study in children exposed to air pollution. For some of these genes, their currently known functions are biologically relevant to environmental or carcinogen exposure (see Table 1 in **Chapter 6**). The focus in this study was on cancer risk in adult humans, with the underlying hypothesis that adults can be exposed long enough to suffer a plausible risk for the development of cancer. The results of this study showed the potential of gene expression to separate regions of inhabitation with differential environmental exposures as gene expressions and the profile as a whole were observed to significantly differ between regions with different characteristics. The most significant differences were observed between the regions of Olen (presence of non-ferro industrial company) on the one hand and the regions of fruit cultivation and the city of Gent on the other. Furthermore, gene expressions demonstrated significant associations with DNA damage as measured by means of COMET assay and with serum protein levels of tumor markers. This chapter therefore concludes that gene expression profiling is a promising tool for application in human biomonitoring studies which analyze environmental carcinogenic risk.

Sensitivity and specificity are important features for a biomarker. The studies presented in the present thesis all returned considerable numbers of significantly modulated gene expressions in exposed compared to non-exposed samples or individuals. The results of the studies on human populations were analyzed at group level throughout the thesis and only (the most) statistically significant results were reported, decreasing the potential influence of inter-individual variability. It has not been established how to assess sensitivity of gene expression as a biomarker, because gene expression profiling generates multiple endpoints. Sensitivity of altered gene expression is a sensitive matter and can be interpreted in various ways; with regard to the magnitude of the effects measured (fold up- or down-regulation), to the levels of significance of the difference between exposed and non-exposed groups or to the possibility to reflect low dose exposures or small differences in exposure.

Sensitivity might be interpreted by comparing the magnitudes of the effects in expression (fold-change) with those observed for the classical biomarkers, such as DNA adducts, micronuclei and COMET data. For example, in the study on cigarette smoke-induced differential gene expressions in monozygotic twins (**Chapter 3**), the mean DNA adduct level was 1.67 fold higher in smokers than in non-smokers whereas the most significantly modulated and real-time PCR confirmed gene expressions showed 1.11 to 1.45 fold differences (on a log 2 scale 0.25 to 0.54). In addition, in the study on Czech children (**Chapter 5**), the over-expression of micronuclei in the Teplice children was 1.38 fold as compared to the Prachatice children ($p=0.023$). Among the 26 most significantly differentially expressed genes ($p<0.0001$ for the difference between Teplice and Prachatice), the magnitude of the effects was between 0.19 and 0.52 on a log2 scale; 1.14 to 1.43 fold. From the biomonitoring study presented in **Chapter 6**, it was observed

that the majority of gene expressions significantly differed between two or more regions of inhabitation, whereas this did not count for all of the classical biomarkers or these markers did not differ as significantly as the gene expressions. In addition, the magnitudes of the differences between regions displaying the lowest and the highest values for the gene expressions and the other parameters were observed to be quite similar; ranging from 1.2 (*DGAT2*) to 2.0 (*ATF4*) and from 1.10 (COMET count) to 2.43 (COMET median) respectively. Thus, in the studies presented here, overall, the effects (differences between exposed and non-exposed or between the region with the highest level and the region with the lowest level) on gene expression were comparable to the effects in classical biomarkers. In addition, the study described in **Chapter 5** reports differential gene expressions in populations under daily life exposures to their environment, which is considered a rather low exposure. The numbers of genes observed to differ in expression between individuals from the air polluted and the relatively clean area and the degree of significance of these differences also indicate sensitivity of gene expression analysis.

Specificity of gene expression was *a priori* considered or hypothesized to be high, as the analysis of large numbers of gene expressions simultaneously delivers more generic and biologically relevant information on cellular effects induced by carcinogen exposure than the classical biomarkers such as DNA adducts or micronuclei frequencies encompass. With regard to specificity of gene expression profiles for specific chemicals, gene expression signatures have been described [1-3]. Therefore, a specific set of genes could in theory be developed for response to specific exposures to or mixtures of chemicals, such as cigarette smoke or air pollution. Our findings indeed suggest gene expression profiling to be specific as we identified different expression modulations responsive to cigarette smoking and to air pollution and also identified genes not reported previously. Although the types of environmental carcinogen exposures as investigated in the present thesis, represent the most abundant carcinogenic mixtures for humans in daily life, to further elucidate specificity of gene expression as a biomarker, future research should concentrate on investigation of (populations exposed to) more and other chemical carcinogens.

Other recommendations and future directions are the following.

- Investigate more chemical carcinogens at more concentrations and different exposure durations *in vitro*, to extend knowledge on chemical specific, dose- and time-responsive transcriptomic reactions.
- For the biomonitoring study as described in **Chapter 6**, information on exposure in terms of measurements of internal dose markers to *e.g.* PAH, PCB and heavy metals, should be incorporated to gain better insight into the individual exposures and to associate these internal dose markers with the gene expressions. This will complement the broad spectrum of biomarkers measured in the Flanders adults and provide a unique dataset for purposes of environmental risk assessment.
- Perform additional (transitional) studies among human populations under different exposure conditions measuring gene expressions and multiple classical biomarkers along the continuum as presented in Figure 2 in **Chapter 1**. Linking gene expression responses to these biomarkers is prudent to elucidate the role of altered gene expression in relation to phenotypic cellular changes.

Summary and General Discussion

- Incorporate information on polymorphisms in key genes for genotoxic exposure into the analyses, either at the level of study design, *e.g.* perform analyses of wild-type or mutant individuals, or at the level of data analysis, by dividing the study population according to polymorphism profile, *e.g.* worst case or best case profiles with regard to functional changes caused due to the polymorphisms.

In conclusion, the data presented in the present thesis indicate that gene expression profiling is a valuable tool for application in molecular epidemiology. The presented results open doors for its application in human biomonitoring studies as gene expression profiling encompasses more generic information on the events occurring after carcinogen exposure in addition to the currently used biomarkers. This research has therefore laid a fundament for the future application of toxicogenomics in molecular epidemiology.

References

1. Hamadeh, H.K., Bushel, P.R., Jayadev, S., DiSorbo, O., Bennett, L., Li, L., Tennant, R., Stoll, R., Barrett, J.C., Paules, R.S., Blanchard, K. and Afshari, C.A. (2002). Prediction of compound signature using high density gene expression profiling. *Toxicol Sci* 67, 232-240.
2. Bushel, P.R., Hamadeh, H.K., Bennett, L., Green, J., Ableson, A., Misener, S., Afshari, C.A. and Paules, R.S. (2002). Computational selection of distinct class- and subclass-specific gene expression signatures. *J Biomed Inform* 35, 160-170.
3. Hamadeh, H.K., Bushel, P.R., Jayadev, S., Martin, K., DiSorbo, O., Sieber, S., Bennett, L., Tennant, R., Stoll, R., Barrett, J.C., Blanchard, K., Paules, R.S. and Afshari, C.A. (2002). Gene expression analysis reveals chemical-specific profiles. *Toxicol Sci* 67, 219-231.

Samenvatting

Het onderzoek beschreven in dit proefschrift is uitgevoerd met als voornaamste doel het ontwikkelen van genexpressieprofielen als een nieuwe biomarker van vroege respons op blootstelling aan chemische carcinogenen, voor uiteindelijke toepassing in een humane biomonitoring studie. De verscheidenheid aan biomarkers die momenteel in gebruik zijn in het veld van de moleculaire epidemiologie, omvat nog geen biomarkers die generieke, biologisch relevante en mechanistische informatie bieden. Omdat de respons van een cel op xenobiotische blootstelling, zoals aan chemische carcinogenen, gedeeltelijk het gevolg is van veranderingen in de expressie van genen, wordt het monitoren van veranderde genexpressieprofielen gezien als een veelbelovende benadering voor de identificatie van nieuwe biomarkers. Met het oog op de conceptuele basis van de biomarkers (Figuur 2 in **Hoofdstuk 1**) wordt genexpressie voorgesteld een vroeg biologische respons op xenobiotische blootstelling te zijn, tussen de biomarkers van blootstelling en die van vroeg biologisch effect in, omdat het wordt gezien effecten weer te geven die in principe omkeerbaar zijn.

Ten behoeve van het evalueren van de potentie van genexpressieprofielen als nieuwe biomarker zijn zowel experimentele *in vitro* als *in vivo* validatie en case studies uitgevoerd, voornamelijk gebruik makend van surrogaat doelwit-weefsel, maar ook doelwit-weefsel voor milieublootstellingen. Perifere bloedcellen (PBC) zijn gebruikt voor het grootste deel van het onderzoek omdat dit het algemeen geaccepteerde surrogaat doelwit-weefsel is in de moleculaire epidemiologie ten behoeve van gebruik in humane biomonitoringstudies en een gemakkelijk toegankelijke biologische matrix representeert, die van mensen verkregen kan worden onder weinig invasieve omstandigheden. In die, vaak grootschalige, studies met humane individuen kunnen de doelwitweefsels of cellen, zoals epitheelcellen uit de luchtwegen, niet worden verkregen om ethische redenen.

Het onderzoek is van start gegaan met een exploratieve *in vitro* studie zoals beschreven in **Hoofdstuk 2**, met als doel het onderzoeken van veranderingen in genexpressie in geïsoleerde, rustende, perifere mononucleaire bloedcellen (PBMC). De genexpressierespons van de PBMC op sigarettenrookcondensaat (CSC) of zijn individuele bestanddelen benzoapyreen (BaP), 4-aminobifenyyl (4-ABP), NNK (4-(methylnitrosamino)-1-(3-pyridyl)-1-butanon) en waterstofperoxide (H_2O_2) werd onderzocht gebruik makend van microarrays welke een verscheidenheid aan toxicologisch relevante genen representeren. De blootstellingen induceerden duidelijk significante veranderingen in genexpressie, zowel verhoogde als verlaagde expressie. Deze studie kende twee doelen. Ten eerste werd een set van 16 genen geïdentificeerd welke reageerden op alle carcinogenen en daarom gezien werden een algemene respons op toxische blootstelling te representeren. Echter, wanneer de expressie van deze genen individueel bestudeerd werd vertoonde geen van de genen een consistente verandering in termen van de richting van de regulatie. Daarom werd deze set genen niet gezien als een bruikbare algemene biomarker. Ten tweede werden de meest gevoelige responsen per individueel carcinogeen geïdentificeerd, welke werden gedefinieerd als de genen die (hoog) significante differentiatie regulatie vertoonden bij blootstelling aan de laagste of gemiddelde concentratie. Dit leverde verschillende sets genen op voor elk carcinogeen, waarbij CSC de meeste genen induceerde in vergelijking met de individuele carcinogenen. Dit was in overeenstemming met de *a priori* verwachtingen, omdat CSC een mix is van een groot aantal toxische en carcinogene verbindingen die worden verwacht genexpressie te beïnvloeden. De expressie van een selectie van genen werd geverifieerd met een onafhankelijk

assay, SYBR Green I real-time polymerase chain reaction (PCR), welke de observeringen van de microarray analyses bevestigde. Analyse van veranderde functionaliteit, namelijk overgerepresenteerde Gene Ontology termen voor biologische processen of moleculaire functies in de lijsten van significant reagerende genen, werd uitgevoerd met het online programma Expression Analysis Systematic Explorer (Expressie Analyse Systematische Verkenner; EASE). Immuunrespons en afweerrespons werden geïdentificeerd als meest significant overgerepresenteerde biologische processen. Concluderend heeft deze *in vitro* studie aangetoond dat genexpressieveranderingen ontstaan door blootstelling aan CSC en zijn bestanddelen. Dit resulteerde in sets van gevoelig reagerende genen die een basis kunnen bieden voor de selectie van toekomstige kandidaat biomarker genen, echter welke zullen moeten worden geverifieerd in *in vivo* studies.

De *in vitro* studie werd opgevolgd door de eerste van twee *in vivo* humane biomonitoring studies, welke het onderwerp is van **Hoofdstuk 3**. Perifeer bloed van eenenijge tweelingparen discordant voor het roken van sigaretten, *i.e.* elk tweelingpaar bestond uit een roker en een niet-roker, werd geanalyseerd voor de effecten van roken op genexpressie, wederom gebruik makend van toxicologisch toegepaste microarrays. Eenenijge tweelingen werden bestudeerd om de mogelijke invloed van (verschillen in) genetische achtergrond op genexpressie te beperken. Het PAXgene Blood RNA systeem werd gebruikt, omdat voor deze studie bloed werd afgenomen bij mensen op een locatie waar geen directe of gestandaardiseerde isolatie van PBMC mogelijk was. Opslag en transport van biologische materialen vormt een praktisch punt van aandacht in humane biomonitoring, in het bijzonder voor het huidige onderzoek dat zich concentreert op het instabiele RNA. Het PAXgene systeem biedt RNA stabilisatie en conservering; op die manier wordt het genexpressieprofiel in het bloed *ex vivo* meteen 'bevroren'. Met het systeem werd RNA van hoge kwaliteit geïsoleerd. Deze studie werd uitgevoerd met twee doelen, inclusief een benadering voor het 'benchmarken' of 'fenotypisch verankeren' van de genexpressies door correlaties met een gevestigde en gevalideerde biomarker van biologisch effectieve dosis: aromatische DNA adducten. Ten eerste werden de meest significant differentieel tot expressie komende genen tussen de rokers en niet-rokers geïdentificeerd, door de combinatie van resultaten van verscheidene statistische tests, als die genen met significante resultaten in elk van die tests; *ATF4*, *SOD2* en *MAPK14*. Deze resultaten werden bevestigd door middel van real-time PCR. Daarnaast werd real-time PCR gebruikt voor het analyseren van de expressie van *CYP1B1* en *SERPINB2*, twee genen die geen significantie lieten zien in de microarray analyses. De resultaten toonden significante toegenomen expressie van alle genen in rokers ten opzichte van niet-rokers, wat in overeenstemming is met bevindingen van andere studies (*CYP1B1*) en de voorgaande studie in PBMC blootgesteld aan sigarettenrookcondensaat *in vitro* (*SERPINB2*, **Hoofdstuk 2**). De tweede benadering omvatte het onderzoeken van associaties tussen individuele genexpressies en DNA adduct niveaus. Correlatie analyses onthulden zeven genexpressies die sterk en significant correleerden met de DNA adduct waarden. Wederom werd SYBR Green I real-time PCR gebruikt ter bevestiging van de microarray resultaten. Real-time PCR kon echter de correlaties van genexpressies met DNA adduct waarden niet bevestigen. Daarom werden deze genen niet langer gezien.

De studie zoals beschreven in **Hoofdstuk 4** was erop gericht de effecten van blootstelling aan sigarettenrook te analyseren op het doelwitorgaan. Het doel was het vergelijken van genexpressie responsen in surrogaat en in doelwitweefsel door het onderzoeken van differentiële genexpressies in primaire normale humane bronchiale

epitheelcellen (NHBE) *in vitro* en door de bevindingen te vergelijken met observaties uit de voorafgaande *in vitro* PBMC studie (**Hoofdstuk 2**) en de *in vivo* rokers versus niet-rokers studie (**Hoofdstuk 3**). NHBE cellen werden blootgesteld aan verschillende concentraties CSC gedurende verschillende perioden. Genoomwijde genexpressie werd geanalyseerd met Agilent Human 22k oligonucleotide microarrays, die meer dan 20.000 geannoteerde gensequenties representeren. Een set genen werd geïdentificeerd welke vergelijkbare significante responsen vertoonden op de blootstellingen in zowel surrogaat als doelwitweefsel *in vitro*. Ook werd een aantal genen gezien te reageren in doelwitweefsel alsook in het bloed van rokers *in vivo*. Cytochroom P450 1B1 (*CYP1B1*) was het enige gen dat significante expressieresponsen vertoonde in zowel bloedcellen *in vitro* en *in vivo* als in de NHBE cellen *in vitro*. Naast het onderzoek in verschillende celtypen, zou het gebruik van verschillende typen of platformen van microarrays hiertoe hebben kunnen bijdragen. Verhoogde expressie van *CYP1B1* is in lijn met literatuur over de overexpressie van dit gen na chemische of genotoxische blootstelling in beide celtypen. Concluderend kan gezegd worden dat CSC effecten induceerde in de NHBE cellen en dat deze effecten tot op zekere hoogte de voornaamste effecten zoals eerder geobserveerd in perifere bloedcellen bevestigen.

Een tweede populatiestudie, uitgevoerd in de Tsjechische Republiek, wordt beschreven in **Hoofdstuk 5**. Twee verschillende eigenschappen van deze studie in vergelijking met de eerder beschreven studies, zijn het onderzoek in kinderen in plaats van volwassenen en de chemische blootstelling zijnde luchtverontreiniging. Om het onderzoek naar mogelijke biomarkergenen te verrijken werd deze studie geïncludeerd, daar de blootstelling aan luchtverontreiniging een meer algemene en veelvoorkomende blootstelling aan een complex mengsel van chemische carcinogenen vormt dan sigarettenrook. Kinderen wonend in de omgeving van Teplice in de Tsjechische Republiek werden onderzocht omdat dit gebied bekend staat om zijn (historisch) zeer hoge concentraties van luchtverontreinigende stoffen, zoals (carcinogene) polycyclisch aromatische koolwaterstoffen (PAH). Voor het analyseren van differentiële genexpressies werden Agilent Human 22k oligonucleotide microarrays gebruikt. Genoomwijde genexpressie in bloed van de kinderen uit Teplice werd bestudeerd in vergelijking met dat van kinderen wonend in een relatief schoon gebied in de Tsjechische Republiek, Prachatic. Daarnaast werden de genexpressies onderzocht op associaties met individuele micronuclei frequenties, welke een biomarker voor vroeg biologisch effect representeren. Omdat de blootstelling aan omgevingsvervuiling in deze studie relatief laag is, vergeleken met bijvoorbeeld arbeidsgerelateerde blootstelling aan carcinogenen of het roken van sigaretten, werd geanticiperd op het zien van relatief lage aantallen significant gemoduleerde genexpressies. De analyses waren daarom geconcentreerd op het identificeren van overgerepresenteerde biologische processen en moleculaire functies in de lijsten van significant veranderde genen. Aanzienlijke aantallen genen werden geïdentificeerd waarvan de expressie significant verschilde tussen de twee groepen kinderen. De meest significant verrijkte (overgerepresenteerde) biologische processen in deze lijsten waren gerelateerd aan nucleosoom en chromatine structuur en behoud (opbouw en afbraak), wat deregulatie van de primaire structurele bouwsteen van het DNA suggereert. Ook werd voor veel genexpressies een significante correlatie met micronuclei frequentie gevonden. Uiteindelijk werd in deze studie een set van vijf genen geselecteerd voor bevestiging met real-time PCR: *CXCL1*, *DGAT2*, *TIGD3*, *PINK1* en *FXN*, waarvan alle behalve *FXN* werden bevestigd. Criteria voor selectie waren een statistisch significant

verschil in expressie in Teplice kinderen in vergelijking met de Prachatice kinderen alsook een significante correlatie van de expressie met micronuclei frequentie. Concluderend suggereren de resultaten van deze studie dat genexpressieprofilering een haalbare en veelbelovende toepassing vormt voor toepassing in het onderzoeksveld van milieugezondheid ('environmental health'). Daarom lijkt het relevant meer populaties te monitoren voor genexpressie-responsen op carcinogene contaminanten uit het milieu.

Als laatste beschrijft **Hoofdstuk 6** de toepassing van een genexpressieprofiel, gedistilleerd uit de voorafgaande studies, in humane biomonitoring. Een grootschalige biomonitoring-studie werd uitgevoerd naar differentiële genexpressies in perifere bloed van geselecteerde populaties van volwassenen in Vlaanderen, die verschillende milieu- of omgevings-blootstellingen meemaken als gevolg van hun woonomgeving. Acht regio's met karakteristieke theoretische milieubelasting werden geïncludeerd; steden met druk verkeer, omgevingen van verbrandingsovens of industrieën alsook gebieden met fruitteelt en het platteland. Gezien het benodigde aantal studiedeelnemers in deze grootschalige biomonitoring, de studie omvatte bijna 400 individuen, werd real-time PCR gebruikt voor genexpressieanalyse in plaats van microarray technologie. Ten behoeve van het onderzoek beschreven in dit proefschrift werden microarray analyses uitgevoerd ter ontwikkeling van het genexpressieprofiel zoals het gemonitord werd in de studie beschreven in dit hoofdstuk. Op basis van voorafgaande studies in humane populaties blootgesteld aan sigarettenrook en luchtverontreiniging, zoals beschreven in **Hoofdstukken 3 en 5**, werden acht genen geselecteerd om gemonitord te worden in de Vlaamse populatie: *CYP1B1*, *ATF4*, *MAPK14* en *SOD2* op basis van de resultaten van de studie naar tweelingen discordant voor roken en *CXCL1*, *TIDG3*, *DGAT2* en *PINK1* uit de studie naar kinderen blootgesteld aan luchtverontreiniging. Voor een aantal van deze genen geldt dat hun huidige bekende functies biologisch relevant zijn voor milieu- of carcinogene blootstelling (zie Tabel 1 in **Hoofdstuk 6**). De focus in deze studie was op het kankerrisico in volwassen mensen met als onderliggende hypothese dat volwassenen lang genoeg blootgesteld kunnen zijn om een plausibel risico te lopen op het ontwikkelen van kanker. De resultaten van deze studie toonden de potentie van genexpressie aan in het onderscheiden van woonregio's met verschillende milieublootstellingen, omdat werd waargenomen dat genexpressies en het genexpressieprofiel als geheel significant verschilden tussen regio's met verschillende karakteristieken. De meest significante verschillen werden gezien tussen de regio's van Olen (aanwezigheid van een non-ferro industrieel bedrijf) aan de ene kant en de regio's van fruitteelt en de stad Gent aan de andere kant. Daarnaast werden significante associaties van genexpressies gevonden met DNA schade, zoals gemeten met de COMET assay, en met eiwitconcentraties van tumormarkers in serum. Dit hoofdstuk concludeert daarom dat genexpressieprofilering een veelbelovende toepassing is voor humane biomonitoringstudies die milieuge-relateerde kankerrisico's analyseren.

Sensitiviteit en specificiteit zijn belangrijke karakteristieken voor een biomarker. De studies beschreven in dit proefschrift leverden alle aanzienlijke aantallen significant gemoduleerde genexpressies op in blootgestelde vergeleken met niet blootgestelde cellen of individuen. De resultaten van de studies in humane populaties werden steeds geanalyseerd op groepsniveau en alleen de (meest) statistisch significante resultaten werden gerapporteerd, wat de mogelijke invloed van interindividuele variatie verlaagde. Het staat niet vast hoe de sensitiviteit van genexpressie als biomarker te bepalen, omdat genexpressieprofilering meerdere eindpunten behelst. Sensitiviteit van veranderde gen-

expressie is een gevoelige zaak en kan op meerdere manieren geïnterpreteerd worden; met het oog op de grootte van de gemeten effecten (maat van toegenomen of afgenomen expressie), op het niveau van significantie van het verschil tussen blootgestelde en niet blootgestelde groepen of de mogelijkheid tot het weerspiegelen van kleine verschillen in blootstelling of blootstelling aan lage doses. Sensitiviteit kan onderzocht worden door het vergelijken van de grootte van de effecten in genexpressie (-voud toegenomen of afgenomen expressie) met die gemeten aan de klassieke biomarkers zoals DNA adducten, micronuclei en COMET waarden. Bijvoorbeeld, in de studie naar differentiële genexpressie geïnduceerd door sigarettenrook in eenjarige tweelingen (**Hoofdstuk 3**) was het gemiddelde DNA adductniveau 1.67 maal hoger in rokers dan in niet-rokers, terwijl de meest significant gemoduleerde en met real-time PCR bevestigde genexpressies 1.11 tot 1.45-voudige verschillen vertoonden (op een log₂ schaal 0.15 tot 0.54). Daarnaast was in de studie onder de Tsjechische kinderen (**Hoofdstuk 5**), de overexpressie van micronuclei in de Teplice kinderen 1.38-voud hoger dan in Prachatic kinderen. Van de 26 meest significant differentieel tot expressie komende genen ($p < 0.0001$ voor het verschil tussen Teplice en Prachatic) lag de grootte van het effect tussen 1.14 en 1.43-voud (0.19 tot 0.52 op een log₂ schaal). In de biomonitoringstudie beschreven in **Hoofdstuk 6** werd waargenomen dat de meerderheid van de genexpressies significant verschilde tussen twee of meer woonregio's, terwijl dit niet opging voor alle klassieke biomarkers of deze markers niet significant verschilden in dezelfde mate als de genexpressie. Daarnaast werden redelijk gelijkwaardige grootheden van verschillen geobserveerd tussen de regio's met de laagste en hoogste waarden voor genexpressie en de andere parameters; variërend van 1.2 (*DGAT2*) tot 2.0 (*ATF4*) en van 1.10 (COMET count) tot 2.43 (COMET median) respectievelijk. Dus, in de gepresenteerde studies waren de effecten (verschillen tussen blootgesteld en niet blootgesteld of tussen de regio met de hoogste versus de laagste blootstelling) in genexpressie vergelijkbaar met de effecten in de klassieke biomarkers. De studie beschreven in **Hoofdstuk 5** rapporteert daarnaast differentiële genexpressies in populaties blootgesteld aan hun omgeving in het dagelijkse leven, dat wordt gezien als een redelijk lage-dosis blootstelling. De aantallen genen die werden gezien te verschillen in expressie tussen individuen uit de luchtverontreinigde regio en uit de relatief schone regio en de mate van significantie van deze verschillen verwijzen ook naar sensitiviteit van genexpressieanalyses.

Specificiteit van genexpressie werd *a priori* gezien als hoog, omdat de analyse van grote aantallen genexpressies tegelijkertijd meer generieke en biologisch relevante informatie oplevert over cellulaire effecten veroorzaakt door carcinogene blootstelling dan de klassieke biomarkers zoals DNA adducten of micronuclei frequenties inhouden. Genexpressiesignaturen zijn beschreven met het oog op specificiteit van genexpressieprofielen voor specifieke chemicaliën [1-3]. Daarom zou het in theorie mogelijk moeten zijn een specifieke set genen te ontwikkelen voor de respons op specifieke blootstellingen aan chemische stoffen of mengsels daarvan, zoals sigarettenrook of luchtverontreiniging. De bevindingen suggereren inderdaad dat genexpressieprofilering specifiek is daar verschillende expressieveranderingen werden geïdentificeerd als respons op het roken van sigaretten en op luchtverontreiniging en ook genen werden geïdentificeerd welke niet eerder gerapporteerd zijn. Hoewel de typen omgevings- of milieublootstellingen onderzocht in dit proefschrift de meest voorkomende carcinogene mengsels representeren in het dagelijks leven van de mens, zou toekomstig onderzoek zich moeten

richten op het bestuderen van (populaties blootgesteld aan) meer en andere chemische carcinogenen. Andere aanbevelingen en toekomstperspectieven zijn de volgende.

- Onderzoek naar meer chemische carcinogenen in meer concentraties en verschillende blootsteldsduren *in vitro*, om kennis uit te breiden over chemicaliën-specifieke, concentratie- en tijdsgerelateerde genexpressiereacties.
- In de biomonitoringstudie zoals beschreven in **Hoofdstuk 6** zal het betrekken van meetgegevens over interne dosis markers voor bijvoorbeeld PAH, PCB en zware metalen, beter inzicht geven in de individuele blootstellingen en associaties van deze interne dosis markers met de genexpressies. Dit zal een breed spectrum aan biomarkers gemeten in de Vlaamse volwassenen completeren en een uniek dataset bieden ten behoeve van milieurisicoschatting.
- Uitvoering van genexpressieanalyses in additionele (transitionele) onderzoeken onder humane populaties met verschillende blootstellingomstandigheden alsmede metingen van meerdere klassieke biomarkers uit het continuüm zoals gepresenteerd in Figuur 2 in **Hoofdstuk 1**. Het verbinden van genexpressie-responsen aan deze biomarkers is belangrijk voor het belichten van de rol van veranderde genexpressie in relatie tot fenotypische cellulaire veranderingen.
- Het betrekken van informatie over polymorfismen in genen die belangrijk zijn bij carcinogene blootstelling bij de analyses, vanaf het opzetten van de studie, bijvoorbeeld door analyses te richten op individuen met wild-type of mutant allelen van die genen, of vanaf de data-analyse, door het verdelen van studiepopulaties op basis van polymorfismeprofiel, bijvoorbeeld 'worst case' of 'best case' profielen met betrekking tot functionele veranderingen van het gen door het polymorfisme.

Concluderend kan gezegd worden dat de data gepresenteerd in dit proefschrift erop wijzen dat genexpressieprofielering een waardevolle toepassing vormt voor gebruik in moleculaire epidemiologie. De resultaten openen deuren voor haar toepassing in humane biomonitoringstudies omdat genexpressieprofielering meer generieke informatie behelst over de gebeurtenissen die zich voordoen na carcinogene blootstelling als aanvulling op de momenteel in gebruik zijnde biomarkers. Dit onderzoek heeft daarom een fundament gelegd voor toekomstige toepassing van toxicogenomics in moleculaire epidemiologie.

References

1. Hamadeh, H.K., Bushel, P.R., Jayadev, S., DiSorbo, O., Bennett, L., Li, L., Tennant, R., Stoll, R., Barrett, J.C., Paules, R.S., Blanchard, K. and Afshari, C.A. (2002). Prediction of compound signature using high density gene expression profiling. *Toxicol Sci* 67, 232-240.
2. Bushel, P.R., Hamadeh, H.K., Bennett, L., Green, J., Ableson, A., Misener, S., Afshari, C.A. and Paules, R.S. (2002). Computational selection of distinct class- and subclass-specific gene expression signatures. *J Biomed Inform* 35, 160-170.
3. Hamadeh, H.K., Bushel, P.R., Jayadev, S., Martin, K., DiSorbo, O., Sieber, S., Bennett, L., Tennant, R., Stoll, R., Barrett, J.C., Blanchard, K., Paules, R.S. and Afshari, C.A. (2002). Gene expression analysis reveals chemical-specific profiles. *Toxicol Sci* 67, 219-231.

Dankwoord

Het is af, op dit stukje tekst na. Langs deze weg wil ik allen bedanken die bij hebben gedragen aan het onderzoek en het tot stand komen van dit proefschrift. Een aantal mensen wil ik in het bijzonder bedanken voor hun belangrijke bijdrage.

Prof. dr. J.C.S. Kleinjans, beste Jos, mijn promotor, dank voor je begeleiding, interesse, kritische blik en enthousiasme. Onze besprekingen hebben mij op de juiste momenten de juiste motivatie gegeven.

Dr. J.H.M. van Delft, beste Joost, mijn copromotor, mijn dank voor je email van 25 september 2001, die uiteindelijk resulteerde in onze samenwerking in de vorm van het promotieonderzoek. Ik heb je praktische begeleiding en 'de deur staat altijd open' altijd zeer gewaardeerd en doe dat nog steeds. Met plezier denk ik terug aan onze reis naar de VS in 2003, mijn eerste internationale conferentie.

Jos en Joost, dank jullie voor het gestelde vertrouwen; ik kijk er vol vertrouwen en enthousiasme naar uit om de komende jaren als post-doc met jullie te werken.

Prof. dr. Robert Vlietinck en dr. Marij Gielen, bedankt voor jullie bijdrage aan de tweelingenstudie. Catherine Derom en de medewerkers van het Oost-Vlaams Meerlingen Register, dank jullie hartelijk voor het contacteren, includeren en bemonsteren van de tweelingen.

Many thanks I express to all co-authors of the manuscripts for their expert contributions, the cooperation and critical reviews of the manuscripts.

Many thanks also to Dr. Ilona Jaspers for providing the NHBE cells and the input in Chapter 4 and to Dr. Ivan Rusyn, for the opportunity to visit and get to know an other lab and be in the 'heart' of toxicogenomics, the Research Triangle. It was a truly great experience.

De analisten van GRAT, waarop je altijd kunt vertrouwen, ben ik zeer dankbaar voor hun hulp en adviezen. Ralph, ik dank jou in het bijzonder voor onze samenwerking. Ik vind je een geweldige collega en ben blij dat je straks achter me staat als paranimf.

Simone en Ebienus, dank jullie voor jullie expertise in de wereld van RNA en microarrays waar ik in het begin van het onderzoek veel aan gehad heb en Simone bedankt voor het beantwoorden van mijn vragen over het promoveren. Roger en Ad, ik heb de gesprekken, tips en adviezen erg gewaardeerd. Harald, dank je voor het bijhouden van de microarray data op de GRAT website.

Andere collega's en AiO's van GRAT, dank jullie voor de gezellige tijden tijdens uitjes, avondjes, de koffie of zomaar even op de gang.

Yvon, ook al gaan we steeds verder uit elkaar wonen, je bent een echte vriendin. Ik ben blij dat je als paranimf achter me staat straks.

Familie, schoonfamilie en vrienden, dank voor jullie interesse, het aanhoren van alle verhalen over het onderzoek en AiO-zijn en het (proberen te) begrijpen ervan.

Papa en mama ik ben jullie zeer dankbaar voor wat jullie al mijn hele leven voor me doen en betekenen en ik hoop dat dat nog lang zo mag zijn.

Lieve Erik, wat ben ik blij dat we elkaar ontmoet hebben. Ik ben je dankbaar voor je liefde, vriendschap en je begrip. Dat onze samenwerking voor altijd mag duren!

Bedankt allemaal!

Danitsja

Curriculum vitae

Danitsja Maria van Leeuwen was born in Delft on May 12th 1979. From 1991 to 1997 she attended pre-university education at the Lorentz Lyceum in Eindhoven. In 1997, she moved to Maastricht to start with the study Health Sciences at Maastricht University. She graduated in December 2001 on the sub specialism Biological Health Sciences. In January 2002 she started her PhD project at the Department of Health Risk Analysis and Toxicology at Maastricht University under supervision of Prof. Dr. J.C.S. Kleinjans and Dr. J.H.M. van Delft. The research conducted in the following four years is described in this thesis. In 2004 she visited the laboratory of Dr. I. Rusyn at the University of North Carolina at Chapel Hill in the United States for three months. During her stay in Chapel Hill, she also collaborated with Dr. I. Jaspers from the Centre for Environmental Medicine, Asthma and Lung Biology (CEMALB) at the Environmental Protection Agency Human Studies Facility, which was the basis for the experiments described in Chapter 4 of this thesis. Since February 1st 2006, she is working as a post-doc on the NewGeneris project at the Department of Health Risk Analysis and Toxicology at Maastricht University.

Danitsja Maria van Leeuwen werd geboren in Delft op 12 mei 1979. Van 1991 tot 1997 volgde ze het VWO aan het Lorentz Lyceum in Eindhoven. In 1997 verhuisde ze naar Maastricht om te beginnen met de studie Gezondheidswetenschappen aan de Universiteit Maastricht. In december 2001 behaalde ze haar doctoraaldiploma in de afstudeerrichting Biologische Gezondheidskunde. In januari 2002 begon ze met het promotieonderzoek aan de vakgroep GezondheidRisicoAnalyse en Toxicologie aan de Universiteit Maastricht onder supervisie van Prof. Dr. J.C.S. Kleinjans en Dr. J.H.M. van Delft. Het onderzoek uitgevoerd in de vier jaren daarna is beschreven in dit proefschrift. In 2004 bezocht ze gedurende drie maanden het laboratorium van Dr. I. Rusyn aan de University of North Carolina in Chapel Hill in de Verenigde Staten. Gedurende dit verblijf werkte ze ook samen met Dr. I. Jaspers van het Centre for Environmental Medicine, Asthma and Lung Biology (CEMALB) aan de Environmental Protection Agency Human Studies Facility, waar de basis werd gelegd voor de experimenten beschreven in Hoofdstuk 4 van dit proefschrift. Sinds 1 februari 2006 is ze werkzaam als post-doc op het NewGeneris project aan de vakgroep GezondheidRisicoAnalyse en Toxicologie aan de Universiteit Maastricht.

List of publications and achievements

Publications

- Goossens, V.J., Schreij, G., van der Geest, S., van Leeuwen, D.M., Baas, D.C., Bruggeman, C.A., van der Ven, A.J. (2002). A delay in CD4 cell response after initiation of highly active antiretroviral therapy is associated with the presence of anti-cytomegalovirus but not with anti-herpes simplex virus antibodies. *AIDS* 16(12), 1682-4.
- Proceedings of the PhD student meeting of the Netherlands Society of Toxicology (2005). Edited by D.M. van Leeuwen and A.S. Kienhuis, Maastricht University, Maastricht, the Netherlands. ISBN 90-806058-5-9.
- van Leeuwen, D.M., Gottschalk, R.W., van Herwijnen, M.H., Moonen, E.J., Kleinjans, J.C. and van Delft, J.H. (2005). Differential gene expression in human peripheral blood mononuclear cells induced by cigarette smoke and its constituents. *Toxicol Sci* 86(1), 200-210.
- van Leeuwen, D.M., van Herwijnen, M., Pedersen, M., Knudsen, L., Kirsch-Volders, M., Sram, R., Staal, Y.C.M., Bajak, E., van Delft, J.H.M. and Kleinjans, J.C.S. (2006). Genome-wide differential gene expression in children exposed to air pollution in the Czech Republic. *Mutat Res* 600(1-2), 12-22.
- van Leeuwen, D.M., van Agen, B., Gottschalk, R.W.H., Vlietinck, R., Gielen, M., van Herwijnen, M.H.M., Maas, L., Kleinjans, J.C.S. and van Delft, J.H.M. (2006). Cigarette smoke-induced differential gene expression in blood cells from monozygotic twins. *Carcinogenesis* (*in press*).

Achievements

- Grant from the Universiteitsfonds Limburg/SWOL to attend the 1st Gordon Research Conference on Toxicogenomics, June 22-27 2003, Lewiston, Maine, USA.
- Travel and accomodation grant from the Netherlands Genomics Initiative for a research visit of three months to the University of North Carolina in Chapel Hill, North Carolina, USA.
- Oral presentation at the meeting of the Toxicogenomics Research Consortium (TRC) of the United States, December 7 2004, Chapel Hill, North Carolina, USA.
- Poster award at the meeting of the Toxicogenomics Research Consortium (TRC) of the United States, December 8 2004, Chapel Hill, North Carolina, USA.
- Organization of the 2005 PhD student meeting of the Netherlands Society of Toxicology, June 7 and 8 2005, Organon N.V. Oss, the Netherlands.