Genomics-Based Identification of Molecular Mechanisms behind the Cancer Preventive Action of Phytochemicals: Potential and Challenges

Theo M.C.M. de Kok*, Simone G.J. van Breda and Jacob J. Briedé

Department of Toxicogenomics, Faculty of Health, Medicine and Life Sciences, Maastricht University, 6200 MD Maastricht, The Netherlands

Abstract: High intake of dietary phytochemicals, non-nutritive compounds found in vegetables and fruits, has been associated with a decreased risk of various types of cancer. With the introduction of new “omics” research approaches, technologies providing large scale and holistic data on biological responses to dietary or environmental factors, our understanding of the molecular mechanisms of the preventive action of individual phytochemicals has started to increase rapidly. This understanding contributes to the biological plausibility of the observed link between fruit and vegetable consumption and decreased cancer risk in epidemiological studies. In this mini-review, we present an overview of the characteristics of the different “omics” techniques, with emphasis on transcriptomics, epigenetics, and the analysis of single nucleotide polymorphisms, and evaluate their implications in studies on dietary phytochemicals. We focus particularly on studies in human cell cultures in vitro and in human population studies and discuss the potential and different challenges offered by each technique, as well as future perspectives on applications of these new tools in nutritional genomics research.

Keywords: Antioxidants, dietary intervention, genomics, human cancer prevention, multiomics phytochemicals, synergy, transcriptomics.

INTRODUCTION: CHEMOPREVENTIVE ACTION OF DIETARY PHYTOCHEMICALS

It is widely recognized that dietary habits present a crucial factor in human cancer risk, and that changes in the dietary composition may have a strong impact on tumor initiation and further cancer development [1]. Extensive review of the literature has demonstrated that mostly foods from plant origin are linked with decreased cancer risk [1, 2]. The cancer preventive effects of particularly vegetables and fruits are ascribed to the intake of fiber material and bioactive non-nutrients, also referred to as phytochemicals. These phytochemicals encompass a wide variety of chemical classes, including minerals, amino acids, carotenoids, glucosinolates, flavonoids, dihydrodiol and organosulphur compounds, which may act through influences on multiple pathways involved in the carcinogenic process. As humans may consume several thousands of different phytochemicals every day, it is obvious that establishing causal relationships between general dietary patterns and cancer risk is a complex task. Additionally, several complex interactions amongst dietary constituents or between phytochemicals and genetic factors (gene-diet interactions) may eventually explain the lack of consistency in epidemiological data on cancer preventive effects of dietary factors.

In order to use the full cancer preventive potential of phytochemicals, we need to understand these complex interactions, and have to explore new approaches and methodologies to do so. Where more traditional research designs were successful in establishing relationships between dietary intake and only a limited number of relevant endpoints and molecular mechanisms involved, the use of genomics techniques in functional food research may expand that horizon and allow us to evaluate effects at a genome-wide level.

In this mini-review, we aim to evaluate the potential role of genomics approaches in identifying relevant phytochemicals for cancer prevention. We also explore to what extent the understanding of complex molecular mechanisms brought forward by these approaches can be of use to establish optimal levels and combinations of these micronutrients in the human diet. Apart from reviewing the literature on genomics responses in human studies, with particular focus on gene expression profiles established in either human cells in vitro or in human in vivo research, we report on research designs and data analysis strategies we are currently applying in our ongoing research.

NUTRITIONAL GENOMICS APPROACHES IN CANCER PREVENTION RESEARCH

In recent years, research in carcinogenesis has made remarkable steps forward, to a large extent based on the use of new genomics-based molecular biological research methods. Crucial genetic factors that predispose to specific types of cancers have been identified as well as multiple genetic defects that may be induced by various environmental, dietary or infectious agents. Our knowledge about these important genetic hallmarks of tumorigenesis and the molecular processes involved may provide new molecular targets for cancer prevention, as well as new strategies in cancer prevention.

*Address correspondence to this author at the Department of Toxicogenomics, Faculty of Health, Medicine and Life Sciences, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands; Tel: +31 43 388 1091; Fax: +31 43 388 4146; E-mail: t.dekok@grat.unimaas.nl

© 2012 Bentham Science Publishers
research. Fig. (1) summarizes how the same genomics techniques applied to improve our understanding of the carcinogenic process may also be used to evaluate the cancer chemopreventive potential of phytochemicals. Such nutritional genomics approaches are expected to result in the identification of genomics-based markers (e.g. gene expression or metabolite profiles) that can not only be used for the prediction of potential beneficial health effects, but may also give more insight into the actual mode of action. These biomarkers may include DNA methylation patterns, transcriptomic, proteomic and metabolomics profiles, which all reflect cellular processes responding to the exposure to phytochemicals. Although genetic polymorphisms are not influenced by the diet, the analysis of different single nucleotide polymorphisms (SNPs) may be relevant to determine gene-diet interactions, and to interpret other genomics responses to phytochemical exposures [3, 4].

Genetics: SNP Analysis and Gene-Diet Interactions

Polymorphisms in genes involved in absorption, circulation, or metabolism of phytochemicals, can affect their potential cancer preventive effects [5, 6]. Until recently, only a few SNPs could be analyzed simultaneously using conventional PCR-based techniques and many studies have reported the effects of nutrigenetics in relation to cancer risk [7, 8]. In particular, the interaction between diet and polymorphisms in genes encoding for phase I or II metabolizing enzymes is widely studied [5, 9]. For example, London et al. [9] investigated the relation between urinary excretion of isothiocyanates and lung cancer risk, and found that individuals with detectable isothiocyanates in urine, and bearing the glutathione S-transferase M1 (GSTM1) homozygous deletion were at decreased risk. This effect was even stronger in individuals with both a deletion of GSTM1 and glutathione S-transferase T1 (GSTT1). In another study, polymorphisms of N-acetyltransferase 2 (NAT2) and GSTM1 were examined in volunteers consuming a vegetarian or a high meat diet [7]. DNA damage was measured in exfoliated colorectal mucosal cells. It was found that the high meat diet significantly increased DNA damage in NAT2 rapid acetylators and among individuals with a GSTM1+ genotype, while it showed no increase in the other phenotype groups.

Since the identification of millions of SNPs by the International HapMap Consortium [10] and the introduction of array-based genotyping techniques, it is possible to screen the entire genome for genetic variations, referred to as genome-wide association studies (GWAS) [11]. However, GWAS in studies with a relatively small number of subjects will be limited by statistical power, and therefore the number of SNPs that can be measured at the same time. To circumvent this, a pre-selection of relevant genes can be made, an approach that has been successfully applied in a human dietary intervention study performed at our department [12]. This study showed that polymorphisms in 6 out of 34 selected genes significantly influenced the outcome of the intervention in 168 subjects. Preliminary data on the impact of these SNPs on gene expression will be presented in section entitled “Application of genomics tools in molecular epidemiology”.

Transcriptomics: Analysis of Gene Expression Responses

Early investigations in gene expression analysis already revealed that micronutrients are very capable of changing the

---

**Fig. (1).** Nutritional genomics and the identification of genomics markers for cancer preventive effects; these biomarkers include genetic profiles (single nucleotide polymorphisms), gene expression profiles, DNA-methylation patterns, protein profiles and metabolite profiles. Modified from ref. [4].
expression levels of specific genes, and flavonoids are probably the first and most extensively studied phytochemicals [13].

Methodologies

Several different techniques are available to study gene expression responses, which can be subdivided into two categories of application: i) transcriptomic studies aiming at the identification of new genes, gene sets and complex gene expression profiles that can be interpreted into modulation of specific molecular pathways (whole genome DNA microarrays, serial analysis of gene expression (SAGE)) and ii) studies investigating differential gene expression of a preselected set of interesting genes (low-density, dedicated microarrays). Both DNA microarray techniques and SAGE are used to produce a snapshot of the mRNA population in a sample. Since arrays can contain tens of thousands of probes, a microarray experiment may identify complex transcriptomic responses to phytochemicals. In contrast, SAGE is a sequence-based sampling technique that is not based on hybridization but provides insight in the transcriptome in the form of small tags corresponding to fragments of those transcripts. For SAGE, mRNA sequences do not need to be known a priori, implying that genes or gene variants which are not known can be discovered.

DNA microarrays have been the technology of choice for large-scale studies of gene expression levels, and the ability of these arrays to simultaneously interrogate thousands of transcripts makes this technique very valuable to establish gene expression responses to phytochemicals. Nonetheless, array technology has also several limitations, for example, background levels of hybridization, effects of different hybridization properties of different probes, and arrays are limited to interrogating transcripts with relevant probes on the array [14].

Sequencing-based approaches (RNA-seq) to measuring gene expression levels have the potential to overcome these limitations. It has for instance been demonstrated that the information in a single lane of Illumina sequencing data is comparable to that on a single array in enabling identification of differentially expressed genes, while allowing for additional analyses such as detection of low-expressed genes, alternative splice variants, and novel transcripts [14]. It is expected that RNA-seq will replace microarray-based techniques and SAGE.

The gene expression changes induced by several phytochemicals, including quercetin, genistein, epigallocatechin 3-gallate (EGCG), vitamin C and E or wine polyphenolics such as resveratrol and organosulphur compounds have been investigated in microarray experiments as recently reviewed by Knasmuller et al. [13]. However, all studies presented in this review are performed in cell cultures or animal experiments, but not in human dietary interventions. We will now discuss the challenges in transcriptomics analysis in more detail including findings reported so far in the small number of publications on human in vivo studies.

Challenges in Transcriptomics Analysis

A growing number of studies use transcriptomics techniques to investigate effects of phytochemicals, mostly focusing on effects of single compounds in cultured cells. Although in vitro studies may generate important information on gene expression changes and molecular pathways, the use of different experimental settings, incubations, analytical strategies and isomeric forms of phytochemicals, make the interpretation a true challenge.

Experimental Conditions and Chemical Characteristics of Phytochemicals

Phytochemicals can occur as different (stereo)isomers or conjugated forms which may differ in their kinetic, toxicological and biological properties. For example, Baron et al. [15] showed a difference in gene expression of several genes induced by the 9-cis and all-trans isomers of carotenoids. However, different isomers of vitamin E (synthetic all-rac α-tocopherol, or natural RRR-α-tocopherol) did not induce different gene expression responses [16]. Brand et al. [17] compared enantiomers of the aglycone form of the flavonone hesperidin, hesperitin. In nature, the S-form is predominant, while in vitro experiments are performed with the commercially available racemic mixture. Although significant differences in transport and metabolism, like conjugation, were found, gene expression by the R- and S-form was comparable. In addition, non-conjugated forms of phytochemicals have been extensively studied in vitro, while mainly the conjugated forms are found in food [18]. Quercetin-conjugates display differential antioxidant properties in vitro [19], and the differences between these could be revealed by genomewide transcriptomic analyses. The use of different solvents also needs to be taken into consideration. Most phytochemicals are only soluble in organic solvents like dimethyl sulfoxide (DMSO), ethanol and tetrahydrofuran (THF). To avoid cytotoxicity, maximally 0.5% solvent is used in cell incubations, but comparable ethanol concentrations can change expression of multiple genes [20], indicating that solvents could disturb gene expression induced by the investigated phytochemicals. Furthermore, these solvents are potent radical scavengers and able to interfere with antioxidant effects.

In Vitro Transcriptomics Studies

The advantage of using a single cell type as compared to more complex in vitro systems, is that results are not disturbed or diluted by complex interference by deviant transcriptomic responses produced by different cell types [21]. Gene expression changes induced by phytochemicals may be the consequence of direct radical scavenging activity, induction of antioxidant enzymes, modulation of phase I and II metabolism, repression of inflammatory or induction of anti-inflammatory mechanisms, and interference with other signaling pathways.

The most direct approach to investigate beneficial health effects is by interpretation of gene expression changes induced by a single phytochemical [22]. For example, van Erk et al. [23] reported a broad range of effects caused by quercetin showing the influence on genes involved in cell proliferation, tumor suppression, cell adhesion and signal transduction.

Phytochemicals can act as direct radical scavengers, and detection of the antioxidant activities of phytochemicals in
cultured cells may require the stimulation of reactive oxygen species (ROS) by oxidants, like H₂O₂ or menadione, to a level that causes measurable damage. ROS triggers cell signaling molecules and a key sensor in the early response is the antioxidant responsive element (ARE) binding transcription factor nuclear factor (erythroid-derived 2)-like 2 (NFE2L2 or NRF2) [24]. NRF2 activation induces ARE-mediated genes resulting in an upregulation of antioxidants enzymes. Previously, we showed that different oxidants modified gene expression profiles showing different temporal effects and pathway changes [25], suggesting that also antioxidant responses to phytochemicals should be evaluated using multiple oxidant systems. On the other hand, most cultured cell lines are derived from tumor cells that already display increased levels of ROS [26]. Furthermore, oxygen levels in cell cultures are higher than under normal physiological conditions, so it may not be necessary to apply oxidants.

Biological interpretation is further complicated by the observation that some phytochemicals activate NRF2 responses via ROS formation [27]. Using specific inhibitory antioxidants can discriminate between oxidant-induced genomics responses and other effects of phytochemicals. For example, by the addition of catalase, EGCG and EGCg-derived H₂O₂-induced gene expression can be distinguished [28]. High levels of phytochemicals can result in pro-oxidant effects, like oxidative DNA damage, but this can subsequently activate beneficial repair responses. Treatment with pro-oxidant levels of vitamin C induced acute response genes, but no induction of genes related to cell cycle arrest, DNA repair or apoptosis was found [29].

Finally, it was suggested that the chemopreventive properties of interactions among various dietary ingredients may result in the potentiation of the activities of single constituents, and therefore these interactions may better explain the preventive effects of whole foods and diets in epidemiological studies. In an earlier review, an overview is presented of different mixtures and combinations of polyphenols with other classes of phytochemicals that show synergistic effects [30].

In conclusion, for the correct interpretation of phytochemical-induced transcriptomic effects in cultured cells, studies are needed that compare multiple phytochemicals on well defined cell types using comparable incubations, different doses, time points, and phenotypical endpoints.

**Human Studies Using Transcriptomics**

Only a few human studies exist in which the effects of dietary phytochemicals are established using transcriptomics. Applying a dedicated array with 256 cytokine/human receptor genes, Majewicz et al. [31] showed the potential of transcriptomics analysis to establish effects of micronutrients. In this study, the impact of vitamin C supplementation was determined in monocytes from smokers and non-smokers. In total 22 genes showed a 2 fold change in non-smokers and 71 genes in apoE4 carriers/smokers. In smokers, 4 key genes were differentially expressed: tumor necrosis factor-beta, tumor necrosis factor-receptor, TRK1-transforming tyrosine kinase protein, and monocyte chemotactic protein-1. In a study from our department [32], two groups of adenoma patients and healthy controls received either a 50% decreased (=75 g/day) or doubled (=300 g/day) vegetable intake for 2 weeks. Gene expression analysis in colon biopsies, using dedicated arrays containing 600 relevant genes, showed that mainly genes related to cell cycle control and genes for oxidoreductase activities were differentially expressed. Twenty genes were modulated that were known to be related to (colon)carcinogenesis. Almost all effects of altered vegetable intake were mechanistically linked to cellular processes that explain either prevention of colorectal cancer risk by high vegetable intake or increased colorectal cancer risk by low vegetable intake.

In a third study, a combined mix of selected dietary factors (resveratrol, green tea extract, α-tocopherol, vitamin C, (omega-3) polyunsaturated fatty acids, and tomato extract) were given as supplements to 36 healthy overweight men with mildly elevated plasma C-reactive protein concentrations in a crossover study with treatment periods of 5 weeks [33]. The transcriptomes of peripheral blood mononuclear cells and adipose tissue were quantified using a custom designed genechip containing 23,941 probe sets [34]. Results showed changes in inflammation-, oxidative- and metabolism-related processes.

In another study, 6 weeks selenium supplementation in 16 subjects resulted in subtle gene expression changes in 1256 genes (P<0.05) from lymphocytes, as detected by 22K arrays. Pathway analysis indicated that most genes encoded for proteins active in protein synthesis [35].

The results from human studies offer relevant information about molecular mechanisms modulated by phytochemicals, but also raise questions about the type of cells/tissues that need to be studied.

**Epigenetics and microRNAs**

There is a growing body of evidence that dietary factors can also modify gene expression by epigenetic and post-transcriptional mechanisms. Several different mechanisms controlling gene expression at the transcription level have been identified, including DNA methylation, histone modifications, and chromatin remodeling factors [36]. Additionally, microRNAs (miRNAs) regulate gene expression at the post-transcription level, by targeting messenger RNAs for cleavage or translational repression, eventually resulting in silencing of the target mRNA and thus silencing specific gene expression [37]. Most widely studied among the gene expression control mechanisms are DNA methylation and histone modifications, although interest in effects of dietary factors on miRNAs is increasing.

Changes in DNA methylation have been observed in many different cancer tissues, and global hypomethylation, associated with chromosomal instability, is commonly observed in cancer cells [38]. On the other hand, region-specific hypomethylation has been linked with oncogene overexpression whereas site-specific hypermethylation has often been associated with silencing of tumor suppressor genes [39]. DNA methylation can be influenced by food components through different mechanisms, of which the presence of methyl donors such as folic acid, choline or betaine, is known to be crucial. Further, DNA methylation is catalyzed by DNA methyltransferases (DNMTs) of which
the activity can also be modulated by dietary compounds [40, 41].

Histone modification is another epigenetic mechanism that can be influenced by the diet [41, 42]. Generally, acetylation of histones is associated with an open chromatin structure and active chromatin, whereas methylated histones are associated with gene silencing. The acetylation state is controlled by the balance between the activities of two antagonistic enzyme systems: the histone acetyltransferases (HATs) that transfer acetyl groups to the lysine residues of the histones, and the histone deacetylases (HDACs) that in turn hydrolyze these modifications [43].

Studies on epigenetic changes mainly focus on DNA methylation, and were until recently restricted to either global DNA methylation or the methylation status of specific DNA sequences. Currently, a wide range of methods is available to establish genomic DNA methylation, either based on bisulfite treatment of the DNA, or on DNA immunoprecipitation (MeDIP). The latter technique provides the basis for high-throughput whole genome detection methods such as high-resolution DNA microarrays (MeDIP-chip) or next-generation sequencing (MeDIP-seq) [43]. In addition to MeDIP, chromatin immunoprecipitation provides the basis for identifying the location of modified histones. This technique is currently combined with microarray technology, like Chip-on-chip and Chip-seq. These high-throughput methods generate huge amounts of data and data-analysis is the major challenge in this field.

The fact that epigenetic regulation of gene expression is involved in cancer progression and that epigenetic changes are also reversible, make these processes a promising target for drug development and cancer risk prevention strategies. Epigenetic activities have been reported for several phytochemicals, of which an overview is given in Table 1. The effects of the dietary flavonoids on promoter methylation status and transcriptional activation of methylation-silenced genes has recently been reviewed by Gilbert and Liu [41]. Also other classes of phytochemicals, such as organosulphur compounds and isothiocyanates are known to influence H3 and/or H4 acetylation and to inhibit the activity of the HDAC enzyme [54, 55, 57]. In vivo, a single dose of sulforaphane ingested by humans resulted in an induction of H3 and H4 hyperacetylation associated with HDAC inhibition in peripheral blood mononuclear cells, suggesting that these cells could be used as biomarkers for human HDAC-inhibition studies [58].

Recent evidence suggests that dietary compounds as diverse as EGCG, folate, genestien, retinoids, and curcumin exert cancer-protective effects through modulation of miRNA expression [63, 64]. miRNAs are short noncoding RNAs that are important in post-translational gene regulation, including regulation of cell proliferation, apoptosis, and differentiation processes. miRNAs are involved in cancer initiation and progression and their expression patterns serve

Table 1. Epigenetic Effects of Dietary Phytochemicals

<table>
<thead>
<tr>
<th>Compound</th>
<th>Dietary Source and Class</th>
<th>Epigenetic Mechanism</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGCG</td>
<td>Green tea catechin</td>
<td>Demethylation of hypermethylated CpG islands of tumor suppressor genes</td>
<td>Reactivation of silenced tumor suppressor genes e.g. p16, RARβ, O6-methylguanine transferase, and hMLH1 in different cell lines; particularly after prolonged treatment</td>
</tr>
<tr>
<td>Genistein</td>
<td>Soy isoflavone</td>
<td>Demethylation of tumor suppressor genes; histone acetylation</td>
<td>Reactivation of silenced tumor suppressor genes e.g. p16, RARβ, O6-methylguanine transferase, and hMLH1 in different cell lines; synergistic effect of HDAC and DNMT inhibition. Increased H3/H4 acetylation p16 and p21</td>
</tr>
<tr>
<td>Caffeic acid /</td>
<td>Catechol containing</td>
<td>Inhibition of DNMT</td>
<td>Inhibition of DNA methylation through increased formation of SAH</td>
</tr>
<tr>
<td>Chlorogenic acid</td>
<td>coffee polyphenols</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selenium</td>
<td>Trace element</td>
<td>Modulation of DNA methylation status</td>
<td>Deficiency results in global hypomethylation</td>
</tr>
<tr>
<td>Folate</td>
<td>Dietary methyl donor</td>
<td>Modulation of DNA methylation status</td>
<td>Deficiency results in global hypomethylation; synergistic interaction with selenium</td>
</tr>
<tr>
<td>Sulphoraphane</td>
<td>Isothiocyanate found in</td>
<td>Histone acetylation / HDAC inhibition</td>
<td>Increased Histone H4 acetylation within the p21 promoter both in vitro (prostate, kidney and colon cell lines) and in vivo (Apcmin mice and human subjects)</td>
</tr>
<tr>
<td>Phenethyl</td>
<td>e.g. cruciferous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>isothiocyanate</td>
<td>vegetables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diallyl disulfide</td>
<td>Organosulfur compound</td>
<td>Histone acetylation</td>
<td>Enhanced histone acetylation and selective modification of histone methylation for chromatin remodeling</td>
</tr>
<tr>
<td></td>
<td>found in allium</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References
[39, 44-46]  [39, 47-49]  [50]  [51]  [52, 53]  [54-58]  [59, 60]  [61, 62]
as phenotypic signatures of different cancers. miRNAs may be useful as biomarkers of cancer prevention or nutritional status, as well as serve as potential molecular targets that are influenced by dietary interventions. For instance, Tsang et al. [63] showed the importance of miRNAs in the regulation of the biological activity of EGCG and identified the role of miR-16 in mediating the apoptotic effect of EGCG. Furthermore, curcumin was shown to induce apoptosis through a miRNA pathway involving caspase-10 as a target of miRNA-186* [65].

Proteomics, Metabolomics and Integrative Multiomics

Although all proteins are coded by mRNA, it is not realistic to expect that the presence and activity of enzymes can be accurately predicted based on transcriptome analysis only [66]. As the proteome consists of all proteins present in a specific cell type it may actually provide a better indication of the biological impact of dietary intake as compared to the transcriptome. However, high throughput analysis of the highly variable proteome is still a major challenge. Although proteome analysis may hold great promise for application in studies on cancer preventive effects of dietary phytochemicals, until now it has mainly been applied in in vitro and animal studies, focusing on effects of specific compounds, and a targeted analysis of proteins [13, 67-69]. Also a limited number of human dietary intervention studies with either cruciferous vegetables, vitamin C or α-tocopherol have shown relevant proteomics responses [31, 70-72]. These studies demonstrate that indeed proteomics analysis, particularly when combined with transcriptome analysis, may reveal effects of dietary phytochemicals relevant in cancer prevention.

Metabolomics approaches aim to identify changes in relevant physiological processes, based on modifications in the occurrence and concentrations of all end-products of metabolic enzyme activity in response to exposures or changes in environmental conditions. One of the biggest advantages of this technique is that it can be applied to biological samples like urine, serum or plasma, making it very suitable for biomonitoring purposes. Although metabolomics analysis can bring comprehensive understanding of biological processes a step forward, studies on the effects of dietary phytochemicals are still very sparse. For instance, Solanky et al. [73, 74] described soy-induced alterations in protein, fat and carbohydrate metabolism based on plasma metabolites related to the dietary intervention, whereas Wong et al. [75] indicate metabolic alterations in the urinary metabolite profile after isoflavone consumption. Other studies focused more on the kinetics and metabolism of the phytochemicals themselves. Fardet et al. [76] described the analysis of different types of urinary metabolites of lignans in a study that illustrates that metabolomics allows the identification of new metabolites of phytochemicals and can also be used to distinguish individuals fed different phytochemical-containing foods.

Although each “omics” technique has its own merits and limitations, combined analysis and interpretation of multi-omics data are likely to offer the best opportunities for comprehensive understanding of the biological influences induced by phytochemicals. Data integration of different “omics” techniques will become increasingly important in nutritional genomics studies [77, 78]. Development of integrated data analyses tools is just starting and several approaches and opinions have been published [77]. A promising tool is Bayesian networks analysis, which enables an integrative, multiplex correlation of different “omics” with phenotypic biomarkers [79].

APPLICATION OF GENOMIC TOOLS IN MOLECULAR EPIDEMIOLOGY

As mentioned above, the number of epidemiological studies applying “omics” technologies is very limited, but from these studies we learned that the results are valuable for the understanding of phytochemical-induced effects. Below we will describe the approach undertaken within an ongoing large-scale human dietary intervention study, in which different “omics” technologies are applied to establish the biological pathways associated with phenotypic markers of effect and exposure within different genetic susceptible groups. A detailed description of the study design has been published [12]. In short, 168 healthy volunteers consumed a blueberry/apple juice rich in quercetin, anthocyanins and vitamin C for a period of 4 weeks, in order to evaluate the preventive properties of fruit-borne antioxidants present in a complex mixture. After the intervention, plasma concentrations of quercetin and vitamin C, and total plasma antioxidant capacity were significantly increased. Moreover, 20% protection (P < 0.01) against ex vivo hydrogen peroxide provoked oxidative damage in lymphocytes as was detected by the comet assay. Upon classification of subjects according to 34 genetic polymorphisms, 6 appeared to influence the outcome of the intervention.

In a follow-up of this project, whole genome microarray analysis was performed to study gene expression modulation in relation to markers of antioxidant action that were already established in the group as a whole and in various subgroups showing a protective effect against oxidative DNA damage. We used a paired study design, in which each subject acted as his/her own control. Microarray hybridization and primary data analyses were carried out as described [80]. In addition to the whole study population for which comet data were available (n = 147), subgroups were defined according to the magnitude of the effect in the comet assay, according to sex, and according to polymorphism (GSTT1*0 or XRCC1*4) (Table 2). For each subgroup, gene lists were generated by means of pair wise t-tests (P<0.05). Venn diagrams were generated in which the significantly changed genes were incorporated for the subgroups under one heading, thereby showing the number of genes which were specifically modulated in each subgroup separately (complement) (Table 2), and modulated by all subgroups (intersection). As a number of genes are similarly modulated within all subgroups under one heading, these can not provide information on gene regulation within a specific subgroup. For example, in men (subgroup 6.1) and women (subgroup 6.2) a number of genes will be similarly modulated, but these genes don’t provide any information about sex-specific responses. In contrast, to investigate the contribution of a specific subgroup, e.g. men (subgroup 6.1) to the effect observed in another group, e.g. to all subjects (group 1), similarly modulated genes are of interest. This approach enables us to specifically relate gene
Table 2. Gene Expression Changes in Defined Subgroups as Analyzed by Whole Genome Gene Expression in Lymphocytes of Subjects Consuming One Litre of Blueberry Juice Per Day for 4 Weeks

<table>
<thead>
<tr>
<th>Group</th>
<th>Subgroup</th>
<th>N</th>
<th>Effect on Intervention on Oxidative DNA Damage</th>
<th>Significantly Modulated Genes ∗; Subgroup Specific (Complement Genes in Venn Diagram)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One group</td>
<td>1) All subjects for which comet data were available</td>
<td>147</td>
<td>↓↓</td>
<td>9485</td>
</tr>
<tr>
<td>Two groups</td>
<td>2.1) All subjects mean different tail moment &lt; -2</td>
<td>74</td>
<td>↓↓↓↓</td>
<td>4304</td>
</tr>
<tr>
<td></td>
<td>2.2) All subjects mean different tail moment &gt; -2</td>
<td>73</td>
<td>↑↑</td>
<td>7776</td>
</tr>
<tr>
<td>Three groups</td>
<td>3.1) All subjects mean different tail moment &lt; -2</td>
<td>74</td>
<td>↓↓↓↓</td>
<td>3528</td>
</tr>
<tr>
<td></td>
<td>3.2) All subjects mean different tail moment &gt; -2 and &lt; 2</td>
<td>37</td>
<td>No effect</td>
<td>1099</td>
</tr>
<tr>
<td></td>
<td>3.3) All subjects mean different tail moment &gt; 2</td>
<td>36</td>
<td>↑↑↑</td>
<td>5126</td>
</tr>
</tbody>
</table>

Sex

| Two groups | 6.1) Men | 46 | ↓↓                                                 | 3088 |
|            | 6.2) Women | 101 | ↓                                                  | 4125 |

Polymorphisms

| GSTT1 | 4.1) GSTT1 wildtypes | 122 | ↓↓                                                 | 5765 |
|       | 4.2) GSTT1 homozygous | 25 | ↑                                                  | 2179 |
| XRCC1*4 | 5.1) XRCC1*4 wildtypes | 57 | ↓↓↓↓                                             | 1410 |
|       | 5.2) XRCC1*4 hetero- and homozygous | 91 | ↓                                                  | 6274 |

$ Different groups were analyzed: based on the effect in the comet assay; based on sex; and, based on polymorphisms.
§ Subgroups defined per analyzed group
£ Number of subjects per subgroup.
∗ Arrows represent direction of effect, number of arrows indicate magnitude of effect.
# Include annotated and non-annotated genes. Significantly differentially expressed gene lists after the intervention were generated by means of pair wise t-tests ($P<0.05$). Venn diagrams were created in which the significantly differentially expressed gene lists were incorporated for each group, thereby showing the number of genes which were specifically modulated in each subgroup separately.

expression changes to a particular subgroup with a specific property, in this example men (subgroup 6.1). In addition, also the contribution of the gene expression response from a particular subgroup with specific characteristics to the total effect, in this example the contribution of the effect in men (subgroup 6.1) to the effect observed in the whole group (group 1), can be revealed. This approach increases the power in finding relevant underlying molecular pathways.

Even though the magnitude of effect and the number of significantly modulated genes for the different subgroups is relatively low as compared to pharmaco- and toxicogenomic studies, this is a commonly observed phenomenon [81] and in line with other dietary intervention studies [35, 82, 83].

Although individual gene expression differences can be of interest, interpretation of the intervention effect on biological processes can be studied using pathway analyses tools. At the moment, the complement and intersection gene sets are exported from the Venn diagrams and transferred into these tools. We expect that the number and nature of significantly modulated pathways will vary between different (sub)groups, as the number and type of genes also differed. In addition to common pathways, specific pathways will provide information on the biological processes which are particularly relevant for these (sub)groups.

Correlation analyses between gene expression changes and phenotypic markers will provide information on whether there is a linear relationship with measured parameters. This approach has been successfully applied in the field of toxicogenomics in order to identify chemically induced gene expression changes related to well-defined indices of toxicity [84]. Applying this approach in our study, Spearman’s rank correlations between gene expression and levels of oxidative DNA damage for the different (sub)groups will identify significantly correlating genes that can be included in further pathway analysis.

In conclusion, the application of genomics in molecular epidemiology will offer more insight into the underlying molecular mechanisms, allowing discovery of biomarkers to assess efficacy of chemoprevention and investigation of the relation between exposure and effects. In particular, using phenotypic anchoring and investigating specific subgroups with defined genetic background, will increase the power of
functional interpretation of large data-sets, resulting in an improved understanding of the mode of action of intervention-related effects.

CONCLUDING REMARKS

Phytochemicals have been identified as dietary factors that may have an important impact on cancer risk development. This has been demonstrated in both epidemiological and animal studies, whereas our understanding of the molecular mechanisms behind the cancer preventive action is mainly based on results from in vitro studies. With the introduction of new “omics” techniques, our insight in the mechanisms of action expands rapidly, and it is exactly this insight that adds biological plausibility to the observed associations between high intake of foods rich in phytochemicals and reduced cancer risk. These new developments enable the identification of omics-based biomarkers that can be applied in future molecular epidemiological studies and which are likely to improve their sensitivity. Results from the earliest studies in this field have demonstrated that both transcriptomic and proteomic changes can indeed be established after human dietary interventions with phytochemicals. Gene expression changes have been established in both target tissues and in lymphocytes as surrogate tissue, indicating the possibility of using specific gene expression profiles, preferably in combination with other “omics” markers, in human population studies using biological materials such as blood and urine, from already existing biobanks. This approach is already being applied in a limited number of projects focusing on genomics responses to environmental and dietary exposures [85, 86].

Apart from the possibilities that “omics” techniques offer, also challenges and limitations have been identified. These relate to the application of different study designs and experimental conditions that are applied in in vitro studies, which make the interpretation of the outcomes into potential health effects in humans rather difficult. Also, each technique has its own technical limitations, which emphasizes the fact that the integration of different genomics markers along with phenotypical endpoints is the most promising approach. This Systems Biology approach requires further development of integrative data analysis tools, and although progress is being made in this field, no applications have been reported in relation to studies on dietary phytochemicals.

Genomics research on phytochemicals offers several attractive future perspectives. Apart from the application of genomics markers in epidemiological studies, the evaluation of the impact of genetic polymorphisms on gene expression profiles and other genomics responses may identify susceptible subpopulations and groups that may particularly benefit from dietary interventions with phytochemicals. This opens possibilities to arrive at individual dietary advice that takes genetic variability into account. The selection of genes to be studied and the interaction between different polymorphisms appears to be especially important in the study of variable responses to dietary factors. Therefore, there are still some scientific as well as ethical hurdles to overcome before sound personalized nutritional recommendations can be provided. Another future perspective is the potential use of genomics markers in establishing antagonistic and synergistic effects of combinations of phytochemicals based upon understanding of the combined and complementary modes of action. Particularly in view of the fact that several phytochemicals are known to possess both beneficial and detrimental effects depending on the dose [87], establishing the optimal combinations and intake levels may lead more effective cancer risk prevention strategies. Furthermore, understanding combined modes of action of different classes of phytochemicals may be valuable in selection and cultivation of varieties of crops that eventually contribute to an optimally balanced diet aimed at the prevention of cancer.

REFERENCES

[16] Muller, P.Y.; Netscher, T.; Frank, J.; Stoecklin, E.; Rimbach, G.; Barella, L. Comparative quantification of pharmacodynamic parameters of chiral compounds (RRR- vs. all-rac-alpha tocopherol)


