NUTRIGENOMICS AND NUTRIGENETICS:
THE OMICS-REVOLUTION IN NUTRITIONAL SCIENCE

Running title: the omics-revolution in nutritional science

Edwin C.M. Mariman, PhD,
Professor of Functional Genomics,
Research Institute NUTRIM,
Maastricht University,
Department of Human Biology,
P.O. Box 616, 6200 MD Maastricht,
The Netherlands.

Tel. +31433882896
Fax. +31433670976
Email: e.mariman@hb.unimaas.nl
ABSTRACT

The execution of the Human Genome Project has brought forth a wealth of information about the structure of the genome, which can now be used to study how the interplay between our genes and factors from the environment like nutrition relate to a state of health or disease. To enable such studies novel technologies have been designed in particular to monitor the activity of multiple genes simultaneously on the level of the RNA by transcriptomics, or the level of the proteins by proteomics. In addition, genome information has boosted approaches to study the role of genetic variation to explain individual differences in responses to nutrition, underlying in part the susceptibility for nutrition-related disorders. These new areas of science referred to as nutrigenomics and nutrigenetics, respectively, will increase our fundamental knowledge of the interaction between life processes and our diet or specific components thereof, which may in time lead to the development of novel functional foods to improve the health status of the general population, and to the personalized diet to prevent the onset of nutrition-related disorders in genetically predisposed individuals.
INTRODUCTION
Several decades ago technological developments made it possible to identify mutations in single genes, which interfere dramatically with metabolism and thereby lead to nutrition-related traits and disorders. In addition, the ameliorative and preventive potential of the diet in relation to monogenic disorders was demonstrated, for instance by the treatment of patients with galactosemia and phenylketonuria [1,2]. In this way, the importance of the interaction between nutrition, metabolism and gene expression for homeostasis was more and more recognized. However, at the same time the complexity of this interaction became apparent. Most frequent nutrition-related disorders appeared not to result from the interaction of a single nutrient with a single gene, but of a complex mixture of nutrients with multiple genes. In order to understand this interaction, research was undertaken from both the angle of the diet and of the genes. Nutrition was regarded as readily accessible, but our complement of genes, our genome, with the state of technology of that time was regarded as the great unknown. It was one of the reasons to engage into a worldwide effort to unravel the structure of the human genome: the Human Genome Project (Figure 1).

With the execution of the Human Genome Project approaching its completion, it was realised that a new era in biological and medical sciences was beginning. This is often referred to as the ‘omics’-revolution. New techniques and knowledge from the Human Genome Project were combined with those of established scientific disciplines such as toxicology leading to the term ‘toxicogenomics’, and pharmacology defining the new field of ‘pharmacogenomics’. Accordingly, the introduction of genomics approaches in nutritional sciences lead to the scientific area called ‘nutrigenomics’. With the application of genomics our knowledge about nutrient-gene interaction will increase substantially and this will provide novel entries for the prevention and therapy of complex nutrition-related disorders for instance by the knowledge-based discovery or design of functional foods.

Understanding complex traits and disorders in humans is complicated by genetic variation. It is estimated that when two haploid genomes are compared with respect to DNA sequence, on average 1 in every 500 base pairs differs. This is the result of the long evolution of our genome in which regularly changes were introduced by mutation, which were then subjected to population genetic processes like selection, migration and drift in order to either disappear again or to reach their present day frequency. Most of the genetic variation occurs as polymorphisms in the 90% of our genome that is non-functional and is not reflected in the phenotype. However, variation can also be present in the coding sequences of genes and in sequences regulating gene expression. Usually, the effect of such a polymorphism on the
phenotype is extremely mild and goes unnoticed. Only when a physiological process needs the action of several genes, and all of those genes exhibit genetic variation, then random combinations of gene variants will result in a broad range of slightly different phenotypes of which the extremes may considerably differ. In this respect, genetic variation underlies the difference in susceptibility to nutrition-related traits and disorders. Given the same diet and lifestyle, one person could become obese or develop diabetes whereas the other would not.

Given the above, it is not surprising that a lot of effort is spent on the detection of the genetic variation as the major basis for person-to-person differences in response to diet. Such knowledge may eventually allow the identification by genetic testing of individuals at risk for certain disorders, the onset of which may then be prevented by proper dietary intervention. Understanding how genetic variation influences gene expression and identifying genetic variants as risk factors for human nutrition-dependent disorders is the focus of ‘nutrigenetics’. To this end, specific strategies have been developed, all of which again depend on the recently obtained structural and compositional knowledge of the human genome.

It should be noted that the nutrigenomics and nutrigenetics applications in this review are centered around nutrient-dependent traits and disorders, which is regarded as one of the main areas of interest. However, nutrient-gene interaction covers a much broader field. It includes, for instance, food technology to manipulate perception of satiety in a strategy to prevent overweight, food fermentation by lower organisms in the gut or in the biotechnological setting, and even food composition and quality assessment. This review does not intend to give a complete overview of the field, but rather describe the basic methodology and general developments.

1. NUTRIGENOMICS: MEASURING NUTRITION RESPONSIVE GENOME ACTIVITY

A way to obtain insight in the methodological approaches of nutrigenomics is to see how an experiment is actually set up and performed. The typical nutrigenomics experiment demands a clear plan based on several a priori choices: 1. the actual approach, 2. the model system, 3. the type of nutrition or diet, 4. the appropriate technological methods.

The influence of nutrition on genome activity is studied almost always in a comparative manner either by a direct or an indirect approach. The direct approach involves changes in the nutrients presented to a model system followed by monitoring the changes in gene expression. This includes for instance most of the human intervention studies. The indirect approach involves the study of nutrition-related traits and disorders such as obesity, type 2 diabetes and cardiovascular disorders. In those studies gene expression is compared
between subjects with and without the disorder and from the differences scientists hope to deduce the relevant molecular pathways leading from health to disease under the influence of diet and lifestyle. The results of those studies should lead to new targets for pharmacological or dietary intervention and to novel functional foods.

In nutrigenomics many different model systems are used ranging from *in vitro* cultured cells to animals and humans. Many of our genes are active in a tissue- or organ-specific manner requiring the analysis of gene expression in the relevant biological material. Although methods for safely taking biopsies from humans are rapidly extending and improving, there will always be at least some medical or ethical restrictions. Therefore, scientists often rely on animal models like mouse, rat or pig, although also other models such as the nematode *C. elegans* are increasingly used. The technology of generating transgenic, knock-out and knock-in mice has considerably increased the attractiveness of using animal models. An example is the ApoE3L mouse in which the human ApoE gene has been introduced [3]. The result is, that the ApoE3L mouse has a more human-like lipid profile and is therefore a preferred model system for gene-diet interaction in the context of the metabolic syndrome. In addition, there is a growing number of inbred mutant mouse strains with a monogenic nutrition-related trait which are excellent models for studying the influence of diet on a particular genetic background.

In part of the nutrigenomics studies, human or animal tissue-specific cell lines or primary cells are used. An advantage is that the external influences, i.e. the culture conditions, can be easily controlled or manipulated by changing specific nutrients followed by monitoring of the effect on genome activity. In addition, cell lines are clonal in origin avoiding the complication of genetic variation. Examples of popular cell lines are the human Caco-2 and HT29 cells representing the gut, human HepG2 cells for the liver, and mouse 3T3-L1 (pre)adipocytes representing cells from the fat tissue. It should be noted that cell lines are usually established through immortalization and that cells in culture miss their natural contact with other cell types. Cultured cells therefore have lost part of their original tissue-specific behaviour and the obtained results should be interpreted with care or tested for their relevance in humans. The latter can also be said for results from animal studies. Leptin was discovered as the adipocyte-secreted peptide hormone, which in rodents suppresses the hunger feeling. Experiments in humans however, failed to show an identical effect [4] but an effect on fertility was observed [5]. Many similar findings clearly indicate that results obtained from animal studies cannot always be translated directly to humans. This is why some scientists argue that humans should be the main subject of nutrigenomics studies. Yet, *in vitro* and
animal studies appear to provide important clues for the mechanisms of gene-nutrient interaction. Although leptin may not exert an identical activity in rodents and humans, hunger is associated with the release of signals from fat cells in both species [6].

Before starting a nutrigenomics experiment the intended changes in nutrition should be carefully considered. One can decide to study the effect of a single nutrient by adding it to the diet in a relatively large concentration. Alternatively, the diet can be enriched for a certain class of nutrient. For instance, the effects of high vs low protein or lipid content on gene expression can be examined. Moreover, the class-composition can be influenced to a certain extent. Vegetable oil contains more unsaturated fatty acids than animal fat, whereas the composition with respect to mono- and polyunsaturated fatty acids depends on the type of oil, i.e. sunflower oil, soya oil, olive oil, etc… Of course, changes in nutrition are easily applied to humans. However, animals have their natural preferences and may be reluctant to certain changes in diet, whereas cells are often very sensitive to variation in the culture conditions. Usually, they tolerate only the addition of one or few components to the culture medium in a limited concentration. Notably, changes in nutrition are always implemented on the background of a basal diet or culture medium, which are complex mixtures of nutrients and could interfere with the activity and the effects of the added components.

What is essential for performing a nutrigenomics experiment is a way to measure the expression of many, if not all genes simultaneously under changing nutritional conditions. In other words, what is needed is a high-throughput approach to measure gene expression of approximately 30,000 genes in case of humans. When genes are expressed, the genetic code of the DNA is transferred to mRNA in a process called ‘transcription’ and after a complex pattern of processing the mRNA is used to ‘translate’ the coding information into the amino acid order of the corresponding protein. In principle, gene expression activity can therefore be measured by determining the amount of produced mRNA or protein. In line with the ‘omics’ terminology, high-throughput approaches for gene expression measurement fall into the areas of ‘transcriptomics’ and ‘proteomics’, respectively. The most popular approaches will be discussed below.

1.1. Transcriptomics: mRNA-profiling
Suppose that in the context of obesity we would like to study all genes involved in storage of triglycerides in adipocytes. A simple experiment could be to isolate human adipocytes, incubate them in culture medium with and without added triglycerides, and compare the gene expression profiles from these two conditions. Changes in the profile would point to genes
influenced in their expression and might be important for storage. To this end, RNA is isolated from the cells before and after the exposure and the RNA-concentration is compared on a gene-to-gene basis. The three most frequently used methods are cDNA-AFLP (copy DNA - amplified fragment length polymorphism), SAGE (serial analysis of gene expression) and the DNA microarray. For cDNA-AFLP [7], mRNA is converted into double-stranded cDNA and then digested with two restriction enzymes. After linker ligation and limited PCR-amplification with primers complementary to the linker sequences, a second round of PCR-amplification follows with labeled primers mainly complementary to the linker sequences but with a 5’ end of 2 or 3 randomly chosen nucleotides. Depending on these 5’ end nucleotides, with each primer a different set of cDNA-derived restriction fragments is amplified representing a particular subset of mRNAs. After separating the fragments on a polyacrylamide gel, qualitative and quantitative comparison of band patterns before and after changing the triglyceride concentration of the culture medium leads to the detection of genes of which the mRNA production has changed. Individual genes can then be identified by cloning and sequencing of the fragment. Although this approach is rather easy to perform, the number of visualized cDNAs is limited and because of the relatively poor resolution of the system, one band on the gel might represent more than one mRNA species.

With SAGE [8], the cDNA is processed by restriction digestion, ligation and PCR to generate artificial DNA molecules consisting of 30-100 ‘tags’, pieces of approximately 12 base pairs of cDNA. This length is sufficient to make each tag a unique representative of the gene from which the corresponding mRNA was produced. The DNA molecules of linked tags are then cloned and sequenced. The relative abundance of a particular tag is a measure for the abundance of the corresponding mRNA in the original RNA isolate and as such it is a relative measure for transcriptional activity of that particular gene. In our experiment this method would be used to compare the mRNA constitution of adipocytes before and after exposure to triglycerides and thereby determine which and to what extent different genes respond. Of course, a high-throughput cloning and sequencing facility is required. To have a >90% chance of identifying a transcript expressed at three copies per cell, one would need to sequence 300,000 tags [9].

The most popular tool for RNA-profiling is the DNA-microarray or DNA–(micro)chip, which is generated by spotting a piece of each gene on a solid carrier (Figure 2). While initially denatured cDNAs were spotted, nowadays it becomes more common to use 50-70 bases long synthetic oligonucleotides as gene-specific fragments. A microarray for the analysis of human mRNA samples should contain approximately 30,000
spots to include every gene and indeed this is now technically feasible. Microrarrays of 20,000 spots on a 1 cm² carrier are commercially available and the up-scaling to 10-100x that number of spots is underway. This will allow the analysis of separate exons of genes. In fact, via alternative promoter use and primary RNA processing each gene is thought to give rise to 3-4 different transcripts and the extended type of arrays will allow the separate analysis of most of these transcripts.

The application of DNA-microarrays for RNA profiling is based on hybridization, the selective binding of each mRNA, as cDNA, to its complementary gene fragment. In preparing for the analysis, the mRNA is converted into cDNA and a fluorescent dye is attached to it. Commonly Cy3 and Cy5 are employed, lighting up by irradiation with UV-light at different wavelengths. In the above proposed experiment, the cDNA from adipocytes-mRNA before triglyceride exposure can be labeled with Cy3 and the cDNA after exposure with Cy5. To visualize changes in mRNA production both labelled cDNA samples are mixed in a 1:1 ratio and hybridized to the microarray where each cDNA will bind to the spotted DNA of its own gene. After washing, the microarray is scanned to detect the fluorescent signals. If a gene does not change its mRNA production due to the triglyceride exposure, similar amounts of Cy3 and Cy5 cDNA will be present on the spot of that gene, which would light up in yellow. Green and red are the colors coming from spots with an abundance of cDNA from non-exposed and exposed cells, respectively. Based on fluorescent staining intensities, the observed changes in mRNA production of each active gene can be quantified. Although a very popular application, still some experimental and analytical problems occur, which deserve proper attention such as background correction, normalization, and annotation of genes. The problem of cross-hybridization of homologous genes to the same spotted DNA has become less with the use of the more specific oligonucleotides, but still cannot be disregarded.

Another aspect that should be mentioned is the handling and analysis of the myriad of data on gene expression generated by these applications. Questions arise on how to extract the relevant biological information from this, which has induced an explosive development of bio-informatics and bio-statistics. Standardising the microarray application is seen as a way to deal with some of the above mentioned problems and to assure that results of all nutrigenomics experiments are compatible with storage in a single database with uniform accessibility to scientists. For this the MIAME/Nut (Minimum Information About a Microarray Experiment in Nutrigenomics) rules are being designed (www.mged.org/Workgroups/MIAME/miame.html).
1.2. Proteomics: protein profiling

Although techniques exist to monitor the activity of every gene on the mRNA production level, a lot of effort is spent to look at gene expression at the level of the proteins [10]. An important reason is that alternative RNA processing and post-translational modification enable a gene to produce several proteins differing in physicochemical and functional characteristics. Estimates range from 6 to 10 proteins per gene meaning that the proteome may be ten times more complex than the genome. Another reason for analyzing proteins is the observation that mRNA and protein concentration often do not change in parallel [11]. When cells respond to a change in available nutrients by increasing or decreasing the mRNA production from certain genes, this does not mean that also more or less of the protein is produced. Finally, cells are able to react acutely to a changed environment by enzymatic modification of a number of proteins [12]. Such adaptive measures via post-transcriptional mechanisms can change the protein profile even without altering the mRNA levels. Therefore, it is evident that proper information about the dynamics of gene expression in cells responding to a nutrient should include results from protein profiling.

Due to the diverse molecular characteristics of proteins, no single technique is available that can visualize the complete proteome, but a considerable part of the proteome can be analyzed. Let us again perform the same experiment as above and search for the changes in gene expression in adipocytes before and after exposure to triglycerides. This time the proteins are isolated and subsequently separated from each other. This is readily done by two-dimensional gel-electrophoresis, in which proteins are first separated according to their electrical charge and secondly according to mass (Figure 3) [13]. After staining, proteins appear on the gel as individual spots, of which the staining intensity is used to estimate the relative amount of a protein in the sample. The spot patterns from the cells before and after exposure are compared using software to detect changes in the concentration of individual proteins. By staining each of the two protein samples with a different fluorescent dye in an approach referred to as ‘2D-DIGE’ (2D DIfferential Gel Electrophoresis [14]) it has become possible to mix the samples before the separation and subsequently analyze the spots for fluorescent color ratios much the same as with the DNA microarrays. Whatever approach is used, the outcome of such an experiment is a number of proteins changing in concentration as the result of the exposure of the cells. However, looking at spots does not reveal the identity of the proteins of interest. To this end, mass spectrometry in combination with knowledge of the sequenced genome is used. A protein is treated with a proteolytic enzyme, usually trypsin which cleaves the amino acid chain behind every lysine and arginine. Since every protein has
its own specific, gene-encoded order of amino acids, tryptic digestion results in a protein-specific mixture of fragments. A protein-specific mass spectrum is generated from the mixture showing an m/z peak for each fragment, which is referred to as a ‘mass peptide-fingerprint’. Here, knowledge from the genome comes in. Because the sequence of most genes, at least for humans, is known, the coding part of each gene can be translated in silico into an amino acid chain, which can then be cleaved in silico by trypsin. Calculating the exact mass of each tryptic fragment gives a theoretical peptide fingerprint for every gene in the genome. By comparing the experimental peptide fingerprint with all the theoretical fingerprints, a match will reveal the identity of the protein.

Like any protein separation method, 2D gel-electrophoresis has considerable limitations because very acidic, very basic, very small (<10 kDa) and very large (>120 kDa) proteins cannot be visualized. Also hydrophobic proteins like membrane proteins are difficult to analyze by this system. Finally, only the most abundant proteins can be detected. Therefore, other complementary protein separation techniques are applied and constantly improved, which are based on chromatography. In any case the final identification step will make use of mass spectrometry. Peptide fingerprinting is not always successful because of unfavorable physicochemical characteristics of the protein fragments leading to absence of mass peaks in the spectrum. Then, tandem mass spectrometry is used to actually determine the amino acid sequence of protein fragments [15]. Advanced mass spectrometry also allows a better quantification of proteins and the identification of posttranslational modifications [16]. Further, it can give information on protein turnover, on complex formation, etc…. The demands for detailed protein information lead to a regular release of novel and renewed analytical instruments in this field in which sensitivity, capacity and throughput are relevant factors. Designing a nutri-proteomics experiment requires that the scientist makes an intellectual choice about the protein characteristics that will be pursued, about the most optimal analytical methods and about the proper instruments that can give the expected result.

One of the established methods for quantifying proteins is ELISA, the enzyme-linked immuno-sorbent assay. In general, antibodies coated on a solid carrier are used to pull down the target protein from a sample mixture, and a second antibody is used to stain the protein in a quantitative way. Since proteomics aims at the analysis of many proteins simultaneously, arrays of spotted antibodies have become available as a new tool for protein profiling. Many such arrays focus on plasma proteins like cytokines and adipokines, which are protein-classes of interest for nutrigenomics studies [17]. It is to be expected that this technology will develop in the coming years both in performance and in the number of proteins that can be analyzed.
simultaneously. Especially for proteins with a transient expression like transcription factors, antibody arrays are being released for comparative studies [18]. The analysis follows the same principle as with DNA microarrays, i.e. labeling of proteins before and after the exposure experiment with Cy3 (green) and Cy5 (red), respectively, and after incubation and binding to the antibodies the screening of the arrays with proper software to determine the green/red ratio of every spot. To our experience, this application faces the same requirements as for DNA microarrays like background subtraction, proper normalization, etc…

Recently, another promising tool has been launched, the protein microarray [19]. It contains more than 4000 proteins of yeast representing a considerable part of the yeast proteome. Soon, such arrays can be expected for higher species including humans. By this technology other types of information can be collected as for instance on protein-protein interaction and complex formation [20]. It may lead to the generation of so-called protein interaction maps showing which proteins work together as partners in the complex structures of cells and how these interactions change under certain conditions [21,22]. In addition, incubation of such arrays with nutrients may reveal which components of the diet can function as signaling molecules and to which target proteins they bind. It will add again another dimension to our knowledge on nutrient-gene interactions.

In line with MIAME/Nut, The Proteomics Standardization Initiative (PSI) of the HUman Proteomics Organization (HUPO) is one of the initiatives to assure that proteomics data are handled, stored and processed worldwide in a standardized way to create the optimal basis for acquiring novel information.

1.3. Metabolomics: metabolite profiling.
Traditionally, nutrients have been regarded as (precursors of) metabolites and metabolic analyses were routinely performed in nutritional experiments. Although metabolites are in fact not the product of a gene, their existence and concentration depends on the activity of multiple genes. Therefore metabolomics, i.e. the large scale analysis of multiple metabolite concentrations under changing nutritional conditions, is also regarded as being part of nutrigenomics. Metabolomics may be especially suited to assess exposure to nutrients, for measuring compliancy during a dietary intervention, or for determining the bioavailability of nutrients. As for proteomics, high-throughput quantitative analysis of the metabolome requires expensive equipment including gas-chromatographs linked to mass spectrometers and nuclear magnetic resonance instruments [23]. Different technologies are applied for the profiling of different classes of metabolites like lipids, carbohydrates and amino acids.
Entering the field of metabolomics is particularly challenging because the complexity increases dramatically in going from the transcriptome via the proteome to the metabolome. Nevertheless, metabolite-profiling will be of growing relevance for the nutrigenomics field in the coming years.

1.4. Aspects of nutrition:
The diet has long been regarded as a complex mixture of natural substances, which supplies both the energy and building blocks to develop and sustain the organism. However, in recent years our view on nutrition has changed and we have learned that nutrients have a variety of additional biological activities: nutrition as a mixture of bioactive components. Some nutrients have been found to act as radical scavengers and as such are involved in protection against diseases like diabetes and cancer. Other nutrients have shown to be potent signaling molecules acting as nutritional hormones. A classical example is their function as ligands for nuclear receptors [24]. When fatty acids and in particular unsaturated fatty acids reach the (pre-)adipose cell, they are believed to activate the nuclear receptor PPARg, which then migrates to the nucleus. There it binds to the promoter of several target genes like that for fatty acid binding protein, which are thereby transcriptionally activated or de-activated. This molecular event boosts the fat-storing machinery. Interestingly, it was observed that the cellular response is much increased by the extra addition of retinoids, which bind to the receptor RXR. The activated PPARg and RXR can form a complex which is more potent in gene regulation than the single components [25].

The observation, that different nutrients can influence molecular pathways in a synergistic or antagonistic way, has revived the interest for using in nutrigenomics studies mixtures of nutrients derived from natural substances rather than single nutrients. In those experiments, selection of the nutrients is often based on empirical knowledge, a part of which comes from traditional oriental medicine. An example is the study of the beneficial effects of tea ascribed to the presence of anti-oxidants like flavonoids [26]. On the other hand, the practical consequences of using complex mixtures of natural substances in nutrigenomics studies are obvious. First, it is very difficult, if not impossible, to change the relative concentrations of the components, although the fermentation process of tea can be managed in such a way that the concentrations of flavones and flavonoles change. Secondly, in vitro experiments cannot easily be performed, because cell cultures are often too sensitive for fluctuations in the culture conditions brought about by the addition of extra components. Most importantly, it cannot be determined to what extent each individual component of the mixture
contributes to the observed effect making proper improvements of diet very difficult. Despite these limitations the use of such complex mixtures of nutrients taken from natural sources is regarded as highly relevant for the development of nutrigenomics-based functional foods.

1.5. Benefits from nutrigenomics.

An important question is what benefits can be expected from this new scientific discipline. Below some of the major deliverables are listed with respect to nutrition-related disorders:

1. **New biomarkers for nutrition-related diseases.** The comparison of biopsy material between patients and matched controls will reveal genes that are specifically up- or down-regulated during disease progression. Some of those genes or their combined expression profile can be used as biomarkers for different stages of disease. They can be incorporated in diagnostic protocols for determining the right moment for disease stage-specific nutritional intervention or therapy.

2. **Biomarkers to monitor the efficacy of nutritional intervention.** Similar comparative studies before, during and after an intervention will reveal genes and expression profiles indicative of the progress and the success of treatment by nutrition.

3. **Genes and molecular pathways as targets for prevention.** Knowledge about the genes, the molecular pathways that they are part of, and their specific role in the pathogenesis will bring forth novel strategies for prevention of disease or disease progress.

4. **Knowledge-based functional foods.** Trying to normalize a pathogenic gene expression profile can be pursued by using mixtures of specific nutrients as nutraceuticals. Because nutrition can be employed in a natural way, it is expected that the general public will accept it more easily than pharmaca for the actual prevention of disease. In the pre-symptomatic phase expression profiles serving as early biomarkers as well as knowledge on the molecular pathogenesis will be of importance for designing and applying novel functional foods.

As indicated above, nutrients as nutraceuticals are favored to play a role in the pre-disease state, either for optimizing health or for preventing disease, whereas pharmaca are intended to cure manifest disease [27]. In fact, the possibility to use nutrition for the prevention of disease in at risk individuals diagnosed by early changes in their gene expression profile or by their genetic constitution, has raised high expectations for (near) future developments within the food industry. Some companies which already supply a number of functional foods are not only using nutrigenomics to search for new targets, but also to broaden the health claim of
products that are already on the market. For instance, a product which has a beneficial influence on the serum lipid profile, can be studied by DNA microarray analysis for its effect on the expression of genes in well-selected target tissues. Suppose it turns out that genes involved in free radical scavenging or immune function are up-regulated. This new information could then be employed to enlarge the health claim and thereby broaden the market for the product. Of course, expected beneficial effects should be demonstrated by properly designed intervention studies.

2. NUTRIGENETICS: A MATTER OF POPULATION VARIATION

As mentioned in the introduction, human genes all have their variants in the population and since response to nutrition is a multigenic process, it is not surprising that non-related individuals may respond differently. In the context of nutrition-related disorders genetic variation underlies variation in risk for disease, in the response to treatment or intervention, and in the chance for successful prevention. On top of knowing which genes are involved, nutrigenetics tries to explain how and to what extent nutrition-related traits and disorders are influenced by genetic variation. In other words, the basal questions of nutrigenetics are: which genes are involved in a trait, what is the functional identity of the variation by which people differ for the trait, and how can this knowledge be used for the benefit of the population.

2.1. A nutrigenetics search for genes:

Traditionally, researchers define ‘candidate genes’ for a trait or disorder and search in those genes for variation. This is referred to as the ‘hypothesis-driven’ approach, because the selection of the genes is based on their (suspected) function like the genes for lipoproteins as candidates for obesity and the genes for insulin signaling as candidates for diabetes. In addition to a biochemical or physiological basis, the orthologous genes underlying animal models with a Mendelian segregation of a trait are attractive candidates for the same trait in humans. A well-known example is the db/db mouse, a model for diabetes, from which the gene for the leptin receptor was identified. The human homologue was then tested for involvement in persons with an impaired glucose tolerance [28].

The testing of a candidate gene for its involvement in a trait follows an established protocol, which begins with the search for sequence variation followed by association or linkage studies. A gene is associated with a trait if one variant (allele) has a significantly higher frequency among persons with the trait than in the general population. Other methods rely on the preferred segregation of an allele from parents to an affected child (transmission
disequilibrium test) or to more affected siblings (sib-pair analysis). If a gene is found to be linked to the trait, it remains to be proven whether the associated or linked allele itself is the risk factor, or whether it behaves merely as a marker for a nearby causative variation, with which it is in linkage-disequilibrium. If the genetic variation corresponds with an amino acid variation in the protein, sometimes a functional test is available and the allelic proteins can be compared for their enzymatic activity, DNA-binding affinity, etc… [29]. But when the variation occurs in the regulatory sequences for gene expression, a biological effect would result from a difference in protein production between the allelic forms of a gene. Although proof of the functional relevance of a gene variant in this case may be difficult to obtain, the associated allele can nevertheless be used as a marker to identify the carriers of increased risk for the trait.

Another approach to search for risk-conferring genes for a certain trait or disorder is the total genome scan, a ‘wholistic’ approach. Hundreds of polymorphisms are selected with a random distribution across the genome but with a known chromosomal location. For every polymorphic site it is then determined with the above methods of association or linkage whether an allele is related to the trait. In this way several chromosomal locations, loci, across the genome can be found with at least one risk-contributing gene per locus. Using this approach risk loci have been described for many human nutrition-related disorders like obesity and diabetes [30,31]. Remarkably, studying a trait in different populations often leads to only partially identical loci. It indicates that the risk for a trait is defined by a limited set of common genes with an additional set of population-specific genes. This can in part explain the inter-individual differences in risk across a geographic region. The wholistic approach can also be applied to animal models. A backcross with a trait-free strain will enable the segregation of the contributing genes into different lines of offspring encompassed in a small piece of genetic material from the trait-carrying parent. By determining the chromosomal location of this parental DNA, the risk locus can be traced. This method is called the ‘quantitative trait locus mapping’ [32].

After finding a risk-locus the remaining task is to identify the actual risk-conferring gene and the causative variation. A problem here is, that each locus may still cover millions of base pairs and therefore a considerable number of genes. To find the actual risk-gene would require a considerable effort of functional testing as described above. To assist the identification of disease genes from genetic loci, a ‘haplotype map’ is being constructed for the human genome consisting of a dense network of polymorphisms [33]. The alleles of neighboring sites have a reasonable chance of being in linkage disequilibrium, constituting a
haplotype. Searching for shared haplotypes among patients with a nutrition-related disorder will allow narrowing of the risk loci to a region of only a few genes.

To improve the chance of identifying risk genes nutrigenetics and nutrigenomics may be combined. Nutrigenomics experiments can provide information on which genes are involved in disease progression. If one of those genes is located at a risk locus, it can be regarded as a candidate gene. There is a reasonable chance that variation in this gene is conferring the risk to carriers. The task that still remains is the identification of polymorphisms in the gene and the functional analysis of the different alleles.

2.2. Personalized diet
Genetic experiments in various populations and the investigation of animal and in vitro model systems have already lead to the identification of several hundreds of genes related to specific nutrition-related traits and disorders like obesity and diabetes [34]. Eventually, the nutrigenomics and nutrigenetics research should lead to the complete understanding of the etiology of those traits and disorders with respect to the interacting genes, the dietary components, and the relative risk conveyed by genetic variation in combination with diet. This combined knowledge will allow the genetic profiling of every individual and thereby assess her/his risk for acquiring nutrition-related disorders. Based on this personal risk profile of genetic factors, a ‘personalized diet’ could then be advised by which the onset of a disorder could be prevented or at least delayed. Already nowadays some enterprises offer genetic testing for alleles that were found to be associated with certain traits, like the ApoE allele as a risk factor for cardiovascular disorders and an allele of the alcohol dehydrogenase gene in connection to alcohol sensitivity and (ab)use. However, it is important to note that carriehership for a certain allele may be said to increase the risk for a disorder by 100%, while in absolute terms the risk would usually still be very small like an increase from 0.001 to 0.002. In addition, testing of one genetic factor might give an incomplete or even a wrong view of the situation. For some persons the risk increased by one factor may be compensated for by another, protective factor leading to the wrong conclusion and advice. At the moment, we have no idea of all the genetic factors and how they interact. Till that time advice based on simple testing probably has to remain simple as well and can only sound like “drink less alcohol” or “eat less saturated fat”. In fact, for this sort of advice genetic testing is not needed.

In the near future an alternative for genetic testing could be the use of nutrigenomics expression profiles on the level of the mRNA, the protein or the metabolites. As demonstrated for the mouse, polymorphicity of genes is reflected in the 2D proteome as mass, charge or
dosage differences [35]. This is probably also true for humans. Recently we have been able to cluster individuals with a high or low fat-burning capacity, as identified by the ventilated hood method, by principal component analysis of their 2D proteome from a fat tissue biopsy (Figure 4).

2.3. Neural tube defects: a classical example with a lesson.
In the decades after the second world war, it was observed in many Western countries that the incidences of spina bifida and its developmental complement, anencephaly, gradually decreased. Improvement of the diet was supposed to have brought forth this effect. A large epidemiologic study was undertaken which eventually showed that the enrichment of the diet with vitamins was able to prevent the occurrence and recurrence of spina bifida in humans with more than 50% [36]. Folic acid, or folate, was identified as the active component and now in many countries pre-conceptional folic acid supplementation is advised as a general preventive measure.

At the same time it has become clear that regarding the risk for a child with spina bifida, the women in the population can be roughly divided into three groups: [a] those who are not at risk even at a low folate intake, [b] those who are at risk at low dietary folate but can be helped by increased folate intake, and [c] those who are at risk despite extra folate intake. Genetic predisposition was suspected and candidate genes taken from the folate metabolism were screened for genetic variation associated with risk. This lead to the detection of the 667C->T and 1298A->C alleles in the gene for MTHFR as risk factors [37,38]. These risk alleles are present to some extent in all three groups of women illustrating that genetic testing in the context of a personalized diet would be inadequate. Moreover, not all of the relative risk of the folate responsive women can be explained by these alleles. Therefore, the search for genetic variation in other genes is continued. New candidate genes may be provided by a nutrigenomics search for folate-responsive genes. Often these studies will be performed with animal models since tissues have to be examined that are not readily available from humans.

The studies of spina bifida have shown something else. In one of the models, the curly tail mouse (ct), it was observed that supplementation of inositol was able to prevent the birth of affected litters [39]. The gene for ct was therefore supposed to play a role in the inositol metabolism. One attempt with inositol supplementation in humans has been successful but from this single case no clear conclusion can be drawn about the general effectiveness of this nutrient [40].
mutated in ct [41]. Surprisingly, there is no obvious link between this gene and the inositol metabolism. An explanation could be that the physiologic disadvantages due to the mutated gene can be overcome by stimulating unrelated but compensatory processes with inositol. For instance, both genes might influence the mitotic index of the cells of the developing tailbud by independent pathways. The important message here is that the hypothesis-drive selection of candidate genes is not a guarantee for success.

3. NETWORKING IN NUTRIGENOMICS AND NUTRIGENETICS.

Due to the technological revolution in research on gene expression the nutritional scientific society in the coming time will generate an enormous set of novel data and knowledge. One of the major issues here is the streamlining of the data-acquisition, data-handling and data–analysis to assure the high quality and comparability of data obtained at different sites. The European Community has financed a Network of Excellence with the acronym ‘NuGO’ (www.nugo.org) to start a major effort in this area. Keywords are ‘joint research’, ‘integration’, ‘standardisation’, ‘education’, etc… It is expected that other regional networks will follow and join together to form a worldwide web of nutrigenomics and nutrigenetics.

References.
Figure 1: The omics-revolution in nutritional science.
Figure 2: DNA microarray application

RNA from exposed cells

RNA from non-exposed cells

Cy3-cDNA

Cy5-cDNA

MIX AND HYBRIDISE

Cy3=Cy5

Cy3<Cy5

Cy3>Cy5
Figure 3: Protein profiling by 2D gelelectrophoresis.
Figure 4: Principle Component Analysis of the fat biopsy proteome of 10 persons with a high or low fat burning capacity.