Review

Annals of Nutrition & Metabolism

Ann Nutr Metab 2005;49:355–365
DOI: 10.1159/000088315

Published online: September 21, 2005

Nutrigenomics: The Impact of Biomics Technology on Nutrition Research

Irène Corthésy-Theulaz[a] Johan T. den Dunnen[b] Pascal Ferré[g]
Jan M.W. Geurts[c] Michael Müller[d] Nico van Belzen[e] Ben van Ommen[f]

[a]Centre de Recherche Nestlé, Département de Nutrition, Lausanne, Switzerland; [b]Leiden Genome Technology Center, Human and Clinical Genetics, LUMC, Leiden, The Netherlands; [c]Corporate Research, Friesland Foods, Deventer, The Netherlands; [d]Division of Human Nutrition, Wageningen University, Wageningen, The Netherlands; [e]ILSI Europe, Brussels, Belgium; [f]TNO – Nutrition and Food Research, Zeist, The Netherlands; [g]INSERM Unit 465, Université Pierre et Marie Curie, Paris, France

Key Words
Genomics · Transcriptomics · Proteomics · Metabolomics · Systems biology

Abstract
The interaction between the human body and nutrition is an extremely complex process involving multi-organ physiology with molecular mechanisms on all levels of regulation (genes, gene expression, proteins, metabolites). Only with the recent technology push have nutritional scientists been able to address this complexity. Both the challenges and promises that are offered by the merge of 'biomics' technologies and mechanistic nutrition research are huge, but will eventually evolve in a new nutrition research concept: nutritional systems biology. This review describes the principles and technologies involved in this merge. Using nutrition research examples, including gene expression modulation by carbohydrates and fatty acids, this review discusses applications as well as limitations of genomics, transcriptomics, proteomics, metabolomics, and systems biology. Furthermore, reference is made to gene polymorphisms that underlie individual differences in nutrient utilization, resulting in, e.g., different susceptibility to develop obesity.

Copyright © 2005 S. Karger AG, Basel/ILSI Europe, Brussels

Introduction

Nutrients have long been considered merely as fuel or cofactors. Subsequently, it was appreciated that they can themselves participate in the regulation of metabolic pathways, for instance through the allosteric regulation of specific enzymes, or by modulating hormone secretion. When molecular biology reached laboratories dealing with the physiology of nutrition and metabolism, it was realized that nutrients either directly or through their hormonal consequences were able to strongly modify the expression of genes (and thus of proteins). Also, the influence of variations in the genome on nutrient utilization and diet-dependent health status (obesity, type 2 diabetes, cancer, cardiovascular diseases) was demonstrated.

The use of biomics research to analyze the influence of nutrients on human health is thus based on two observations: (1) the nutritional environment modifies the expression of genes, and (2) depending upon the genotype of an individual, the metabolism of nutrients may vary and ultimately result in a different health status.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2005 S. Karger AG, Basel/ILSI Europe, Brussels
0250–6807/05/0496–0355$22.00/0

Accessible online at:
www.karger.com/anm

ILSI Europe A.I.S.B.L.
Av. E. Mounier 83, Box 6
B–1200 Brussels (Belgium)
Tel. +32 2 771 00 14, Fax +32 2 762 00 44, E-Mail publications@ilsieurope.be

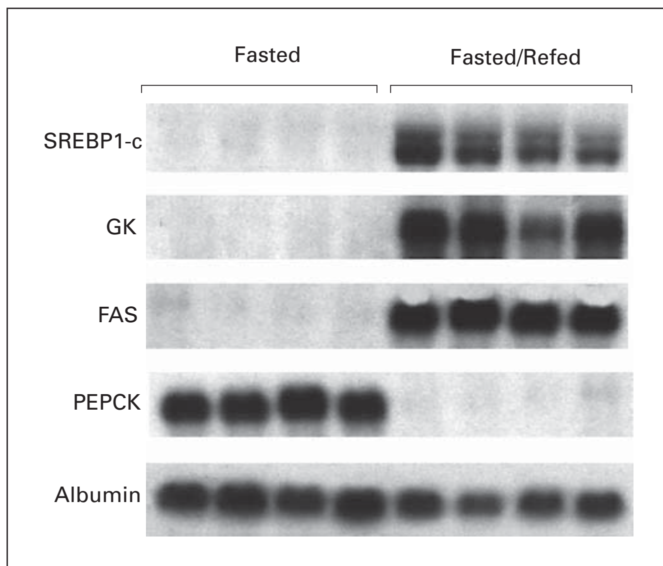

Fig. 1. Liver gene expression in fasted rats and rats refed with a high-carbohydrate diet. The figure shows the mRNA concentrations, detected by Northern blotting, of the genes coding for the transcription factor SREBP1-c, which mediates insulin action; glucokinase (GK), the first enzyme involved in glucose metabolism in hepatocytes; fatty acid synthase (FAS), which converts an excess of glucose into lipids; and phosphoenolpyruvate carboxykinase (PEPCK), which is a key enzyme of de novo glucose production and albumin. Albumin mRNA serves as a control for equal loading of gel lanes.

Nutrients Modulate Gene Expression

The adaptations of energy metabolism to the quality and quantity of food imply the modulation and/or emergence of metabolic pathways. Many of these adaptations indicate a change in the amount of a given protein. This is usually achieved through a change of the transcription rate of the corresponding gene. It has now been demonstrated that major (carbohydrates, fatty acids, amino acids) or minor (e.g. Fe, vitamins) dietary constituents participate in the regulation of gene expression in response to nutritional changes. Nutrients modulate the activity of transcription factors, or the secretion of hormones that in turn interfere with a transcription factor (transcription factors are proteins that bind to specific DNA sequences located within the promoter region of genes and can activate or inhibit their transcription). Most of our knowledge on the effects of nutrients on gene expression has been acquired in animal models. Below are two examples of gene regulation by carbohydrates and lipids, which are characteristic of the transcriptional mechanisms involved. More examples can be found in a recent overview of the mechanisms by which nutrients interact with their molecular targets to modify gene expression [1]. Another technology that, besides animal models, might soon bring important insights in the function of genes is RNAi [2]. Using RNAi, expression of any gene can be easily modulated (reduced) facilitating the direct analysis of its effects. In addition, when a selective screen can be designed, RNAi technology can be very powerful to identify other genes active in the same biochemical pathways that are able to regulate particular processes [3, 4].

A Carbohydrate Rich-Diet Modifies Hepatic Gene Expression

Glucose is used continuously at a high rate in mammals by organs such as the brain (120 g/day in humans), red blood cells, or renal medulla. When a meal that contains carbohydrate is absorbed, it induces several metabolic events aimed at decreasing endogenous glucose production by the liver and increasing glucose uptake and storage in the form of glycogen. If glucose is delivered into the portal vein in large quantities and once the hepatic glycogen stores are repleted, glucose carbons can be converted in the liver into lipids (lipogenesis), which are exported as VLDL and ultimately stored as triglycerides in adipose tissue. Conversely, if glucose availability in the diet is reduced, glucose-utilizing pathways are inhibited and glucose-producing pathways are activated. On the genomic level, these physiologic mechanisms are facilitated by changes in gene expression levels. A high-carbohydrate diet induces the expression of several key glycolytic and lipogenic enzymes in the liver [5]. Figure 1 illustrates the changes in the expression of some hepatic genes involved in glucose metabolism during a fasting-refeeding transition in rats.

Absorption of a carbohydrate-rich diet is concomitant with increases in the concentration of substrates such as glucose, but also with changes in the concentration of pancreatic hormones, insulin and glucagon. It has been shown that some genes such as glucokinase can be induced by a high insulin concentration independently from the presence of glucose [6], whereas others require both an increased insulin and glucose concentration in order to be induced [5]. We will focus here on the mechanisms by which glucose on its own can modulate gene expression in the liver. The reader will find a review on the effect of insulin on gene expression in [5].

Glucose must be metabolized in order to have its transcriptional effect [7]. Glucose-6-phosphate and xylulose-5-phosphate (an intermediate of the non-oxidative branch

of the pentose phosphate pathway) have been proposed as the signal metabolites. Carbohydrate response elements (ChRE), i.e. a short stretch of DNA conferring glucose sensitivity to a gene promoter, have been identified in several glucose-responsive genes [5]. ChRE is bound by the transcription factor ChREBP (ChRE-Binding Protein) [8]. In conditions of glucose shortage, ChREBP is phosphorylated and this induces its sequestration in the cytoplasm. In the presence of high glucose, a phosphatase is activated possibly by xylulose-5-phosphate and the dephosphorylation of ChREBP induces its translocation of to the nucleus and its activation, thereby stimulating the transcription of glucose-responsive genes [9]. Figure 2 summarizes this mechanism.

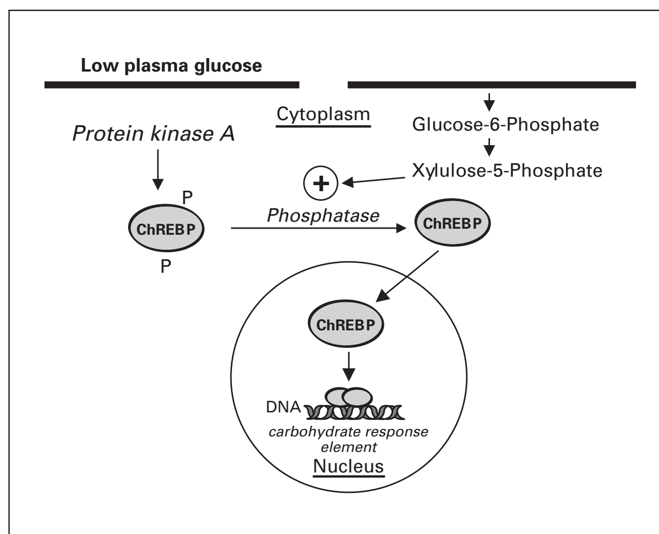

Fig. 2. Mechanism of glucose action on gene transcription in hepatocytes. For explanation, see text. ChREBP = Carbohydrate response element-binding protein.

Fatty Acid-Modulated Transcription Factors

Both high-fat diet and starvation periods are concomitant with a large increase in fatty acid oxidation capacity in various tissues. Part of these adaptations is linked to the effect of fatty acids on a specific transcription factor called PPARα (peroxisome proliferator-activated receptor-α).

In rodents, different amphiphatic acids (such as fibrates) are able to induce a strong hepatic peroxisome proliferation, a proliferative effect absent in humans [10]. The peroxisome proliferation in rodents is concomitant with an increase in the transcription of genes involved in peroxisomal and mitochondrial fatty acid oxidation [10]. PPARα was cloned as a murine nuclear receptor that mediates the effects of peroxisome proliferators [11]. PPARα is highly expressed in hepatocytes, cardiomyocytes, kidney cortex, skeletal muscle (tissues with a high capacity for fatty acid oxidation) and enterocytes [12, 13].

PPARα binds and is activated by unsaturated fatty acids [10, 14]. Saturated fatty acids have a lower affinity for PPARα. In the liver, activation of PPARα by synthetic ligands (e.g. fibrates) induces fatty acid transport proteins and long-chain acyl-CoA synthetase (activation of fatty acids into acyl-CoA is a prerequisite for their subsequent metabolism) [15, 16]. Several key enzymes involved in peroxisomal and mitochondrial ß-oxidation and in ketone body synthesis are also direct targets of PPARα [17–20]. These results have been confirmed by the use of mice in which both copies of the PPARα gene have been inactivated (PPARα-null mice). Whereas the phenotype of normally fed PPARα-null mice is not fundamentally different from wild-type mice, starvation induces major differences [21, 22]. The liver and heart of PPARα-null mice are steatotic ('fatty') owing to a much lower fatty acid oxidation capacity. Circulating ketone bodies are extremely low due to an impairment of hepatic ketone body synthesis, and the mutated mice develop hypoglycemia (utilization of ketone bodies by the brain and of fatty acids and ketone bodies by other tissues allows conservation of glucose in periods of shortage such as starvation or a high-fat diet).

Also in humans, diets rich in carbohydrates or fat modulate gene expression in muscles and adipose tissue [23, 24], demonstrating that these nutrient-driven molecular events are an essential component of our adaptative possibilities.

A potential advantage of the application of transcriptomics (see below) is the identification of common nutrient response elements in the promoter of genes of which the expression is modulated by diet. This could favor the discovery of the transcription factors involved and thus explain the integrated response to a particular signal.

Gene Polymorphisms and Nutrient Metabolism

Phenotypic variability is based on interindividual genetic variation. In contrast to highly inbred murine strains, the human population is characterized by genome diversity due to the presence of many polymorphisms. Polymorphisms can be qualitative, i.e. affect the sequence itself (basepair changes – SNP for single nucleotide poly-

morphisms – small nucleotide deletions, duplications or insertions), or quantitative, i.e. affect the copy number (large duplications or deletions). Qualitative variants can affect the regulatory region of a gene (i.e. the promoter region) or the coding/non-coding sequences; quantitative changes directly affect the level of expression. Although the effective number of common polymorphisms for a single gene in the population is rather small, the combination of alleles in our 30,000 genes is obviously much higher. The inherited genotypic differences in DNA sequence contribute to phenotypic variation and to differences in disease risk in response to the environment. Numerous studies in animals and humans have shown that individual genotypic variations can alter nutrient metabolism, from relatively mild conditions like lactase gene polymorphisms that result in lactose intolerance [25] to potentially severe pathological conditions like phenylketonuria (PKU). One of the most studied examples is the tendency to develop obesity.

Obesity: An Example of the Importance of Genotype in Nutrient Handling

If energy expenses (mechanical work and heat) are smaller than the intake, energy will be stored, mainly as triglycerides contained in adipose tissue. Hydrolysis of stored triglycerides into fatty acids and their subsequent oxidation allows us to face periods of energy shortage or increased energy expenses. Thus, energy storage in the form of fat is an important adaptation for survival. It is therefore likely that during evolution, combinations of genes have been selected that favor fat accumulation (the 'thrifty gene' hypothesis). In the context of increased food availability and decreased physical activity found in the developed world, these genes will confer a susceptibility to the development of obesity and its maintenance [26]. The involvement of genetic factors in the control of body weight is indicated by the high concordance in monozygotic twins of body composition and response to overfeeding [26], and by the discovery of genes involved in monogenic forms of obesity in humans, most of which are part of the central pathways of food intake regulation.

Considerable progress in the understanding of food intake regulation stems from the discovery of leptin. Leptin is a cytokine-like polypeptide produced by adipocytes; it informs the brain of the fat storage level and controls food intake through the activation of hypothalamic leptin receptors [27]. In the hypothalamic arcuate nucleus, leptin induces the synthesis and secretion of α-melanocyte-stimulating hormone (α-MSH) from the precursor pro-opiomelanocortin (POMC) through a pro-hormone convertase 1 (PC-1)-dependent cleavage [27]. α-MSH then inhibits the effectors of food intake. Homozygous carriers of a loss of function mutation in the leptin gene exhibit morbid obesity with onset in the first months of life, hypogonadotropic hypogonadism, and central hypothyroidism [27]. Mutations of the receptor for leptin have also been found [28]. In three subjects homozygous for the mutation, a truncation of the receptor abolished leptin signaling, leading to a phenotype similar to that of individuals with leptin deficiency, although more severe. At present, seven cases have been described for leptin or leptin receptor mutations. Two children with homozygous or compound heterozygous POMC loss-of-function mutations exhibited a phenotype including obesity and adrenal insufficiency, reflecting the lack of pituitary neuropeptides derived from the POMC gene [29]. Mutations of PC-1 have also been described; these are associated with obesity and ACTH insufficiency, as for POMC mutations.

Mutations in the melanocortin 4 receptor (MC4R, the receptor of α-MSH) cause dominant and recessive inherited nonsyndromic obesity with incomplete penetrance (the mutation is not always associated with obesity) and variable expression (a similar mutation can yield various degrees of obesity) [26, 30]. Human obesity caused by MC4R mutations is similar to more common forms of obesity with an earlier age of onset. MC4R mutations represent a significant cause of obesity in morbidly obese children and adults (0.5–6%) [31].

Although these monogenic forms of obesity have been well documented, the genetic susceptibility is in most cases polygenic. Two approaches are possible in order to identify susceptibility genes. The first one consists of genomewide scans in order to detect chromosomal regions showing linkage with obesity in large collections of nuclear families, mostly comprising adult sibling pairs. Fifty-nine chromosomal loci, distributed over all chromosomes except Y, have been linked to obesity; however, no gene involved in common obesity has been characterized up to now [32]. The second method is the candidate gene approach. It involves testing the association between obesity and a specific allele of a candidate gene (e.g. a gene involved in the regulation of food intake), either in a family study or in large cohorts of unrelated controls and patients. Numerous studies have tried to identify 'susceptibility' genes, and more than two hundred genes [32] have been associated with obesity, implying that a large number of polymorphism combinations might exist. However, the relative risk associating a specific gene allele and

obesity remains low because each gene probably contributes only for a small part and it is thus important to consider the association of polymorphisms on different genes.

Obesity is not the only diet-related situation that has been associated with gene susceptibilities. Type 2 diabetes [33], cardiovascular diseases [34] and cancer [35] are also well documented in this respect [36].

Nutrition, Gene Polymorphism and Biomics Studies

It is clear that the quantity and quality of the diet modulates the expression of numerous genes in various tissues. Each person challenged by a diet change might react differently due to the individual combination of gene polymorphisms, even though the same final homeostasis may eventually be reached. Whether global gene expression trends can be detected and used as biomarkers that define 'normality' remains to be proven but constitute a goal for biomics studies in nutrition. Similar questions that could be addressed using biomics are (1) how polymorphisms affect nutrient handling; (2) whether specific combinations of polymorphisms can be associated with diet-related diseases such as obesity, and (3) whether gene profiling can be used to demonstrate that a given nutrient has beneficial or deleterious health effects [1]. Obviously, attempting to answer these questions will be a major challenge.

Biomics Technology

In 1989, the Human Genome Project (HGP) was initiated as a worldwide effort to clone, map and sequence the entire 3.2-Gb human genome in a 15-year period. This enormous task was finished in 2003 with the publication of the complete human sequence with an error rate of less than 1 in 10,000 bases. In the mainstream of the HGP revolutionary new technologies were developed. The main characteristics of these new technologies were miniaturization, automation, high throughput and computerization. Without the developments in the latter area, hardware and software, the HGP would not have been possible.

These technical developments have now given us tools enabling high-throughput 'genome-wide' approaches, i.e. covering, in one analysis, the 'entire human genome'. These tools form the basis of the biomics era; genomics (covering DNA), transcriptomics (RNA), proteomics (protein), metabolomics (metabolites) and systems biology (integrating all of these), with bioinformatics enabling the storage, integration and analysis of the overwhelmingly complex data sets produced.

Genomics

Genomics concerns the analysis of DNA, the genome, focusing on resolving the variation in the genome between individuals. Genomics tries to correlate this variation with all kinds of phenotypic parameters (linkage and association studies). Variations in DNA 'order' (translocations, inversions, transpositions) seem to play only a minor role and cannot yet be analyzed genomewide. Variation in the sequence at the nucleotide level (substitution and small deletions, duplications, insertions) has long been considered most important; variant coding sequences directly translate to variant function and variation can be used for linkage and association studies. New quantitative technologies including array-CGH (comparative genomic hybridization) and quantitative SNP technology revealed an unexpected and large-scale copy number variation (CNV, deletions and duplications) in the human genome [37–39]. The majority of CNVs include genes, thereby directly influencing the expression level (in theory 50% up or down), and should be considered as an important and hitherto neglected type of variation in the human genome. For measurement, SNPs (single nucleotide polymorphisms) are currently the most popular tools, although di-nucleotide repeats (CA repeats) and AFLP (amplified fragment length polymorphisms, mainly employed in plants) are also widely applied. Variation and recombination in the human genome is not spaced randomly and worldwide efforts currently try to get an overview. As a result of these efforts, the so-called HAPMAP project, a minimal set of SNPs should emerge which is sufficient to perform a genome-wide linkage/association study. Currently high-throughput genomewide analysis is facilitated using several technologies, the most powerful being array technology (DNA chips and micro-arrays), mass spectrometry (e.g. MALDI-TOF) and beads-based flow sorting. The maximal capacity lies in the order of 100,000 SNP typings per day. The major problem is not technological but financial, the current cost per typing being around EUR 0.01–0.05.

Transcriptomics

Focusing on the analysis of RNA (the transcriptome), transcriptomics aims at measuring the level of expression of all or a selected subset of genes based on the amount of RNA present in a sample. Currently, the most powerful tool available is DNA array technology. Using one array the expression level of up to 50,000 transcripts can be

measured in parallel, and tens of samples can be screened per day. Examples analysing nutrient biology include the effect of leptin (which mimics a repleted state) on mouse adipose tissue [40]; the effect of fasting and aging in rat hypothalamus, which is responsible for the control of food intake [41, 42]; the effect of diets rich in polyunsaturated fatty acids on murine hepatic and hippocampal gene expression [43]; the effects of anti-oxidants on muscle transcripts in rat muscle [44], muscle gene expression in young and old men [45] and in type 2 diabetic subjects [46], and gene expression-profiling of human visceral adipose tissue [47]. In these studies, hundreds of genes are usually varying in expression. The difficulty is to organize the results in such a way that they can be used to elucidate biological mechanisms, or to derive biological markers for a given physiological situation. Such data treatment is obviously an essential requirement if one wants to understand the overall consequences of nutrient intake.

The main limitation lies in the sensitivity of the assay as well as in data analysis. Statistically significant measurements can only be obtained for the most abundantly expressed genes, and when expression differences are changed by a factor of three or more. When smaller changes need to be detected, the measurement has to be repeated several times, making studies rather costly.

As for any new technology, array-based transcriptomics is hindered by initial limitations in analytical precision and standardization. The standardization issue was noted at an early stage. The Microarray Gene Expression Database (MGED) group was founded in 1999 with the goal of facilitating the adoption of standards for DNA array experiment annotation and data representation, as well as the introduction of standard experimental controls and data normalization methods (www.mged.org). In addition, high per-analysis cost seriously reduces the number of measurements performed per study. Furthermore, different platforms are used, e.g. cDNA vs. oligonucleotide array and micro-array vs. GeneChips, and designs are regularly modified to incorporate new genes and improved probe sequences, thereby complicating data comparison [48]. Several commercial suppliers produce of-the-shelf arrays (e.g. Affymetrix, Agilent, Amersham) or oligonucleotide collections for home-spotting (e.g. Illumina, Operon). Intrinsically, these provide a first 'standardization' which is desperately needed to be able to compare results from different studies.

Because of these problems, studies that adequately meet rigid statistical requirement are in fact relatively scarce [49]. In view of these weaknesses, data from array-based transcriptomics need to be interpreted cautiously. Finally, it should not be underestimated that the major source of variability arises from the starting biological material itself. Ensuring the validity of this material, collecting all possible variables (from genome, to age and environment) and controlling the sampling conditions and timing are essential to obtain meaningful data.

Another limitation lies in the tools used for data analysis (software). Current bioinformatic tools are not yet very effective and opinions vary considerably regarding the best computational algorithms to apply.

Proteomics

Proteomics technology focuses on the analysis of proteins and their interactions, the proteome. The challenge lies in the development of technologies, which are able to cope with the huge differences in chemical properties of proteins as well as the wide dynamic range of protein concentrations. Initially, two-dimensional (2D) gel electrophoresis was used to measure the expression level of a large number of proteins. If the full set of proteins separated by 2D gels is to be identified, automated equipment is used for excising protein spots, digesting the proteins therein, and analyzing the resulting peptides using mass spectrometry. However, 2D-gel analysis is biased towards the most abundant changes, which might lead to erroneous conclusions since also subtle variations may lead to important changes in metabolic pathways. In addition, low abundant proteins and very hydrophobic, acidic or basic proteins are often not detected [50] and identification of the proteins resolved is time consuming and costly. Like with array technology, 2D-gel analysis can be combined with two-color fluorescent labeling, highlighting those proteins that differ in expression between the two samples.

Recently, mass spectrometry (MS) has come into play as an exciting and very powerful analytical tool. Proteins are submitted to proteolytic degradation to form collections of peptides that are subsequently analyzed using MS (peptide mass fingerprints). When the molecular mass of certain peptides does not concur with published structures, these peptides are further characterized by MS-MS tandem mass spectroscopy. Such analyses will also reveal posttranslational modifications of proteins, such as phosphorylation. MS can be used to quickly determine the identity of a specific protein and it facilitates analysis of very complex protein samples, zooming in on those proteins that differ in expression. Impressive studies have been performed in the area of cancer research and diag-

nosis using mass spectrometry in combination with 2D liquid chromatography (2D-LC). The limiting step for this technology lies mainly in data analysis, i.e. computing power and the lack of adequate software tools. Generating MS traces is a matter of seconds with cost of the measurement being a minor issue. Another approach uses isotope coding and subsequent quantification relying on digestion of the protein mixture and separation on peptide rather than on protein level ('shotgun proteomics'). While gel-associated drawbacks are circumvented, isotope coding is limited by the risk of insufficient yields and the alteration of the sample composition. Moreover, it is often compromised by a protein bias due to tagging post-digestion, amino acid-targeted reagents and possible chromatographic separation of the light and heavy labels.

Further developments can be expected from the development of protein-array technology. Using antibody or protein arrays, whole-genome studies can be envisaged which monitor protein expression levels, protein-protein interactions and protein activity (e.g. kinase arrays). Using a combination of DNA and proteins in array technology, e.g. ChIP-on-chip (ChIP = chromosome immune precipitation) experiments should facilitate assays which monitor (genomewide) binding of transcription factors to DNA promoter sequences per se and in response to external stimuli, including environmental factors and nutrients.

Major challenges for the field of proteomics are posed by the study of protein-protein interactions and the relationship protein polymorphisms – nutritional value. To study protein-protein interactions, two-hybrid cloning systems have proven to be efficient and successful techniques [51, 52]. Recently, two-hybrid arrays have been developed, in which the screening is performed in a colony array format with each colony expressing a different pair of proteins [53, 54]. Array screens can be easily automated facilitating high-throughput and reproducible protein-protein interaction screens. Furthermore, comparing the results from several assays circumvents the problem that single assays generate a high number of false positives.

Most biomics questions require the differential analysis of two biological states, typically case vs. control, and, consequently, quantitative 'biomic' tools. It should be noted that while quantitative transcriptomics have evolved into commercialised and partially standardized platforms, quantitative proteomics is just emerging and standards are largely lacking.

Proteins in body fluids like milk are translated from mRNAs that are expressed in different tissues, such as mammary epithelial cells and milk leukocytes. Consequently, milk proteins cannot be readily deduced from transcriptomics, and proteomics has particular relevance here. Applications include identification of minor milk proteins with potent biological function, such as growth factors, and investigation of milk protein polymorphisms. Such polymorphisms have implications for the properties and processing of milk, as well as for its nutritional value [55, 56], which is determined not only by amino acid composition but also by digestibility and digestion rate of proteins. Moreover, polymorphism may change the pattern of peptides released during digestion in the gastrointestinal tract, which may result in differences in biological activity and allergenicity of their peptide mixes. β-Lactoglobulin may be one of the main causative agents of cow's milk allergy in infant, i.e. consumption of milk can lead to an increase in the level of β-lactoglobulin-specific IgG. Consumption of β-lactoglobulin A led to higher levels of antibodies in guinea pigs than the consumption of β-lactoglobulin B; however, allergic response to both variants was similar [57]. This may be related to differences in the tryptic hydrolysis of β-lactoglobulin A and B: tryptic hydrolysis of (partially unfolded) β-lactoglobulin A is approximately three times faster than hydrolysis of β-lactoglobulin B [58], tryptic hydrolysis of β-lactoglobulin B results in larger peptides than β-lactoglobulin A [59].

Metabolomics

Metabolomics technology focuses on the analysis of metabolites, the metabolome. It tries to measure the level of all substances (other than DNA, RNA or protein) present in a sample; the metabolome comprises the complete set of metabolites synthesized by a biological system. Such a system can be defined by level of biological organization, such as organism, organ, tissue, cell, or cell compartment levels. Today, the best tools for metabolomics research are proton nuclear magnetic resonance (NMR) and mass spectrometry (MS). Biologically relevant samples can easily be obtained from blood, sweat, urine, and feces.

Metabolomic analyses have only just begun. NMR-based metabolite profiling through highly quantitative and broad-spectrum classes still suffers from inherent sensitivity issues. The dynamic range is between few nanomoles to few hundred micromoles. Thus most abun-

dant metabolites, i.e. steady-state concentrations are observed. However, recent cryoprobe technology shows great promise to overcome sensitivity hurdles. LC-MS, complementary to NMR, offers superb sensitivity but is limited by the essentially nonquantitative nature of mass spectrometry requiring internal standardization. Addressing automatic spectra processing, a few obstacles are worth mentioning: although peak-recognition software is getting smarter (for NMR and MS), inconsistencies in chemical shifts and baseline shifts in NMR spectra have to be compensated and this is not a standardized task.

Publicly available metabolomic databases should be created similarly to genomic sequence repositories and technical advances as well as improved data mining and analysis tools will need to be developed. Due to pleiotropic effects, the effect of a nutrient may lead to changes of metabolite levels in various, seemingly unrelated biochemical pathways. Therefore, a comprehensive analysis of all metabolites is required to understand such hidden relationships. Both sample preparation and data acquisition must aim at identifying all classes of compounds, assuring high recovery as well as experimental robustness and reproducibility.

Systems Biology: More than the Merge of Biomics Technologies

By helping to understand the interaction between nutrients and molecules in our bodies, the implementation of molecular biology and biochemistry in 'classical nutrition' research, followed by the technological revolution of the 'biomics' technologies, will greatly affect nutritional sciences. The first studies that span the levels of genome, transcriptome, proteome, and metabolome demonstrate this impact. Most of these studies investigate differential effects on the level of (metabolic) pathways, and provide new mechanistic insights. However, the real potential of biomics technologies does not limit itself to such differential display type of strategies, where the measurement of a very large set of parameters is exploited only for those parameters that show eye-catching differences. The complete dataset surely contains much more information.

On top of that, the various 'layers' of the biomics technology platform are of course related (genes encode RNA, which encodes the enzymes that catalyze the conversion of metabolites). Thus, in combining the datasets of genome, transcriptome, proteome and metabolome, a wealth of added information becomes available. In fact, this combination of datasets paves the way to a complete description of the biological behavior of a cellular system, in response to external stimuli.

Although the complexity of this proposed integration (i.e. systems biology) is exceeding the current bioinformatics tools and capacities, its implications for nutritional research can be enormous. Unlike biomedical interventions (drug therapy), nutrition is chronic, constantly varying, and composed of a very large amount of known and unknown bioactive compounds. Furthermore, nutrition touches the core of metabolism by supplying the vast majority of ingredients (both macro- and micronutrients) for maintaining metabolic homeostasis. This homeostasis stretches from gene expression to lipid metabolism and from signaling molecules to enzyme cofactors. Thus, nutrition by its nature *needs* to be studied in an integrated way. So far, most of the tools for this integration were lacking, thus maintaining an unbridgeable gap between classical nutrition (studying human physiology with excursions into biochemical pathways) and biomedical sciences (elucidation disease-related molecular mechanisms). In applying systems biology to nutritional sciences, these paradoxal extremes are bridged and the complexity of the relationship between nutrition and health can be met by the complexity of the integrated approach. At the moment this is little more than a dream, since only premature and pragmatic example studies of this concept are currently being performed. Many hurdles need to be taken, most of them in the field of bioinformatics, before this area matures.

The Systems Biology Biomarker Concept

In the relation between nutrition and health (unlike the relation between nutrition and disease) it is necessary to develop a new concept of biomarker. It needs to reflect subtle changes in homeostasis and the efforts of the body (cellular systems, organs, and interorgan interactions) to maintain this homeostasis. Also, it preferably should include a wide variety of biological actions. Furthermore, both efficacy and safety aspects should be monitored simultaneously. Single nutrients may have multiple known and unknown biochemical targets and physiological actions, which may not be easily addressed with classical biomarkers (i.e. the 'single-gene, protein or metabolite' approach, usually at non-physiological conditions). In addition, the efficacy assessment of health effects of nutritional components is even further complicated by the fact that single dietary constituents are hardly consumed as separate entities but are part of a dietary mixture.

The inter-individual variation on the genetic level, as discussed above, appears to be adding even more com-

plexity to the nutrigenomics picture. Quite a number of these genetic polymorphisms have been described from a clinical genetics point of view, simply because they predispose to a pathological condition. These usually are the monogenic forms, with a pronounced effect on functionality. Given the high number of SNPs present in our genome (>3 million), it is obvious that many will be identified that could effect the relationship between nutrition and health. Instead of pursuing pathological leads, in particular the effects of these 'minor' variants needs to be studied. Establishing their impact on health, in relation to nutrition, is achieved through cohort-type studies. However, such approaches will fail simply due to a loss of power when multiple minor genetic polymorphisms are involved, all acting on the same physiological process. Nutrition research should not ignore these minor variants, because several subtle changes may together produce large effects, e.g. obesity can both be the result of one pronounced SNP as well as an interplay of many less pronounced variants. Nutri-biomics may contribute in this effort, by describing the mechanisms in terms of kinetic and dynamic models, and verification and validation of these models with genotyping combined with functional analysis on the level of RNA, protein and metabolites.

Bioinformatics and Data Mining

Bioinformatics is the technology enabling the data processing, clustering, dynamics, integration and storage of the overwhelmingly complex data sets produced by the biomics research.

Bioinformatics will play a crucial role to condense the massive amounts of data generated through high-throughput experimental procedures used in biomics, and to integrate these with data obtained from traditional techniques. The challenge is to combine all pieces of information so that all data can be looked at in a coherent way. Novel algorithms, software and hardware are being developed to translate sets of gene, protein and metabolite data into biochemical pathways.

In order to gain full access to these emerging powerful tools, it is paramount to address the enormous challenge of unifying complex and dissimilar data [60]. The incorporation of observations from numerous sources and domains into a unified, seamlessly searchable database and turning in it into knowledge will impact every facet of modern nutrition [61]. Two important features are required in order to integrate data between databases; they have to speak the same language, and use the same identifiers for the same object. In other words, the biological domain should be described thoroughly, using controlled vocabularies and ontologies (see http://www.ebi.ac.uk/GOA/project.html) to allow meaningful data comparisons. The gene ontology consortium has developed a dynamic, structured, and precisely defined vocabulary for describing the roles of genes and their products in any organism. The goal of this effort is to address the deficits of the current rather divergent nomenclature schemes [62, 63]. The South African National Bioinformatics Institute and Electric Genetics have developed a set of vocabularies that can be applied to describe the source of samples and materials used in studying human gene expression. This controlled vocabulary was released to the scientific community under an Open Source License [64]. The use of dissimilar identifiers also limits the application of bioinformatics; thus, the establishment of a comprehensive gene index is of prime importance. One of the best known providers of a human genome gene index and gene annotation, the Ensembl project (www.ensembl.org) is an entirely Open Source project and has been widely adopted by academic and commercial organizations [65].

Major challenges that remain to be addressed are to define 'normal' and 'healthy' versus 'unhealthy' profiles, especially in the predisease stage. Sampling location and timing will have to be optimized within ethical and practical constraints. In outbred populations like humans, detection of subtle effects may require numbers of participants that are too large to be realistic, unless participants can be preselected based on their genotype.

Conclusions

State-of-the-art biomics technologies are used to study the effect of dietary habits on health promotion and disease prevention. Not all individuals respond identically to dietary interventions, and meaningful biological interpretation of the generated data is a very complex issue. Intense collaboration between biologists, analytical scientists, statisticians and bioinformaticians is essential in order to gain the most from biomics data. To generate interpretable results one must start with a clear and solid question. The amount of data available in public or proprietary databases is increasing rapidly, underlining the central role of bioinformatics to transform raw data into relevant biological knowledge. Furthermore, use of multiple experimental tools and methods will increase the reliability of the results.

Advancement of research in computational and analytical science will gradually transform nutrition into a more systematic and hypothesis-driven science. To accelerate and coordinate successful application of high-throughput technologies in the field of nutritional research, collaborative actions and networks will be necessary. A good first example is the European Nutrigenomics Organisation NuGO (www.nugo.org), combining the technology and scientific expertise of 22 European academia and research institutes in the area of nutrigenomics, with 'network of excellence' by the European Union.

Acknowledgement

This work was commissioned by the Nutrition and Genetics Task Force of the European branch of the International Life Sciences Institute (ILSI Europe). Industry members of this task force are Coca-Cola Europe, Eurasia and Middle East, Friesland Foods, Groupe Danone, Nestlé, Numico and Unilever. For further information about ILSI Europe, call +32 2 771 00 14 or E-Mail info@ilsieurope.be. The opinions expressed herein are those of the authors and do not necessarily represent the views of ILSI and ILSI Europe.

References

1 Muller M, Kersten S: Nutrigenomics: goals and strategies. Nat Rev Genet 2003;4:315–322.
2 Brummelkamp TR, Bernards R, Agami R: A system for stable expression of short interfering RNAs in mammalian cells. Science 2002;296:550–553.
3 Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, et al: A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature 2004;428:431–437.
4 Paddison PJ, Silva JM, Conklin DS, Schlabach M, Li M, Aruleba S, Balija V, O'Shaughnessy A, Gnoj L, et al: A resource for large-scale RNA-interference-based screens in mammals. Nature 2004;428:427–431.
5 Foufelle F, Ferre P: New perspectives in the regulation of hepatic glycolytic and lipogenic genes by insulin and glucose: a role for the transcription factor sterol regulatory element binding protein-1c. Biochem J 2002;366:377–391.
6 Iynedjian PB, Jotterand D, Nouspikel T, Asfari M, Pilot PR: Transcriptional induction of glucokinase gene by insulin in cultured liver cells and its repression by the glucagon-cAMP system. J Biol Chem 1989;264:21824–21829.
7 Foufelle F, Girard J, Ferre P: Regulation of lipogenic enzyme expression by glucose in liver and adipose tissue: a review of the potential cellular and molecular mechanisms. Adv Enzyme Regul 1996;36:199–226.
8 Yamashita H, Takenoshita M, Sakurai M, Bruick RK, Henzel WJ, Shillinglaw W, Arnot D, Uyeda K: A glucose-responsive transcription factor that regulates carbohydrate metabolism in the liver. Proc Natl Acad Sci USA 2001;98:9116–9121.
9 Kawaguchi T, Takenoshita M, Kabashima T, Uyeda K: Glucose and cAMP regulate the L-type pyruvate kinase gene by phosphorylation/dephosphorylation of the carbohydrate response element binding protein. Proc Natl Acad Sci USA 2001;98:13710–13715.
10 Kliewer SA, Xu HE, Lambert MH, Willson TM: Peroxisome proliferator-activated receptors: from genes to physiology. Rec Prog Horm Res 2001;56:239–263.
11 Issemann I, Green S: Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. Nature 1990;347:645–650.
12 Braissant O, Foufelle F, Scotto C, Dauca M, Wahli W: Differential expression of peroxisome proliferator-activated receptors (PPARs): tissue distribution of PPAR-alpha, -beta, and -gamma in the adult rat. Endocrinology 1996;137:354–366.
13 Su J, Simmons C, Wisely B, Ellis B, Winegar D: Use of a PPAR gamma-specific monoclonal antibody to demonstrate thiazolidinediones induce PPAR gamma receptor expression I. Hybridoma 1998;17:47–53.
14 Desvergne B, Wahli W: Peroxisome proliferator-activated receptors: nuclear control of metabolism. Endocr Rev 1999;20:649–688.
15 Motojima K, Passilly P, Peters JM, Gonzalez FJ, Latruffe N: Expression of putative fatty acid transporter genes are regulated by peroxisome proliferator-activated receptor alpha and gamma activators in a tissue- and inducer-specific manner. J Biol Chem 1998;273:16710–16714.
16 Schoonjans K, Watanabe M, Suzuki H, Mahfoudi A, Krey G, Wahli W, Grimaldi P, Staels B, Yamamoto T, Auwerx J: Induction of the acyl-coenzyme A synthetase gene by fibrates and fatty acids is mediated by a peroxisome proliferator response element in the C promoter. J Biol Chem 1995;270:19269–19276.
17 Brady PS, Marine KA, Brady LJ, Ramsay RR: Co-ordinate induction of hepatic mitochondrial and peroxisomal carnitine acyltransferase synthesis by diet and drugs. Biochem J 1989;260:93–100.
18 Dreyer C, Krey G, Keller H, Givel F, Helftenbein G, Wahli W: Control of the peroxisomal beta-oxidation pathway by a novel family of nuclear hormone receptors. Cell 1992;68:879–887.
19 Gulick T, Cresci S, Caira T, Moore DD, Kelly DP: The peroxisome proliferator-activated receptor regulates mitochondrial fatty acid oxidative enzyme gene expression. Proc Natl Acad Sci USA 1994;91:11012–11016.
20 Rodriguez JC, Gil-Gomez G, Hegardt FG, Haro D: Peroxisome proliferator-activated receptor mediates induction of the mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase gene by fatty acids. J Biol Chem 1994;269:18767–18772.
21 Kersten S, Seydoux J, Peters JM, Gonzalez FJ, Desvergne B, Wahli W: Peroxisome proliferator-activated receptor alpha mediates the adaptive response to fasting. J Clin Invest 1999;103:1489–1498.
22 Leone TC, Weinheimer CJ, Kelly DP: A critical role for the peroxisome proliferator-activated receptor alpha (PPARalpha) in the cellular fasting response: the PPARalpha-null mouse as a model of fatty acid oxidation disorders. Proc Natl Acad Sci USA 1999;96:7473–7478.
23 Cameron-Smith D, Burke LM, Angus DJ, Tunstall RJ, Cox GR, Bonen A, Hawley JA, Hargreaves M: A short-term, high-fat diet up-regulates lipid metabolism and gene expression in human skeletal muscle. Am J Clin Nutr 2003;77:313–318.
24 Vidon C, Boucher P, Cachefo A, Peroni O, Diraison F, Beylot M: Effects of isoenergetic high-carbohydrate compared with high-fat diets on human cholesterol synthesis and expression of key regulatory genes of cholesterol metabolism. Am J Clin Nutr 2001;73:878–884.
25 Harvey CB, Hollox EJ, Poulter M, Wang Y, Rossi M, Auricchio S, Iqbal TH, Cooper BT, Barton R, et al: Lactase haplotype frequencies in Caucasians: association with the lactase persistence/non-persistence polymorphism. Ann Hum Genet 1998;62:215–223.
26 Barsh GS, Farooqi IS, O'Rahilly S: Genetics of body-weight regulation. Nature 2000;404:644–651.

27 Schwartz MW, Woods SC, Porte D Jr, Seeley RJ, Baskin DG: Central nervous system control of food intake. Nature 2000;404:661–671.
28 Clement K, Vaisse C, Lahlou N, Cabrol S, Pelloux V, Cassuto D, Gourmelen M, Dina C, Chambaz J, et al: A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. Nature 1998;392:398–401.
29 Krude H, Gruters A: Implications of proopiomelanocortin (POMC) mutations in humans: the POMC deficiency syndrome. Trends Endocrinol Metab 2000;11:15–22.
30 Vaisse C, Clement K, Durand E, Hercberg S, Guy-Grand B, Froguel P: Melanocortin-4 receptor mutations are a frequent and heterogeneous cause of morbid obesity. J Clin Invest 2000;106:253–262.
31 Dubern B, Clement K, Pelloux V, Froguel P, Girardet JP, Guy-Grand B, Tounian P: Mutational analysis of melanocortin-4 receptor, agouti-related protein, and alpha-melanocyte-stimulating hormone genes in severely obese children. J Pediatr 2001;139:204–209.
32 Rankinen T, Perusse L, Weisnagel SJ, Snyder EE, Chagnon YC, Bouchard C: The human obesity gene map: the 2001 update. Obes Res 2002;10:196–243.
33 Froguel P, Velho G: Genetic determinants of type 2 diabetes. Rec Prog Horm Res 2001;56:91–105.
34 Herrmann SM, Paul M: Studying genotype-phenotype relationships: cardiovascular disease as an example. J Mol Med 2002;80:282–289.
35 Vineis P: Diet, genetic susceptibility and carcinogenesis. Publ Health Nutr 2001;4:485–491.
36 Willett WC: Balancing life-style and genomics research for disease prevention. Science 2002;296:695–698.
37 Fredman D, White SJ, Potter S, Eichler EE, Den Dunnen JT, Brookes AJ: Complex SNP-related sequence variation in segmental genome duplications. Nat Genet 2004;36:861–866.
38 Iafrate AJ, Feuk L, Rivera, MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: Detection of large-scale variation in the human genome. Nat Genet 2004;36:949–951.
39 Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, et al: Large-scale copy number polymorphism in the human genome. Science 2004;305:525–528.
40 Soukas A, Cohen P, Socci ND, Friedman JM: Leptin-specific patterns of gene expression in white adipose tissue. Genes Dev 2000;14:963–980.
41 Kappeler L, Gourdji D, Zizzari P, Bluet-Pajot MT, Epelbaum J: Age-associated changes in hypothalamic and pituitary neuroendocrine gene expression in the rat. J Neuroendocrinol 2003;15:592–601.
42 Li JY, Lescure PA, Misek DE, Lai YM, Chai BX, Kuick R, Thompson RC, Demo RM, Kurnit DM, et al: Food deprivation-induced expression of minoxidil sulfotransferase in the hypothalamus uncovered by microarray analysis. J Biol Chem 2002;277:9069–9076.
43 Berger A, Mutch DM, Bruce GJ, Roberts MA: Dietary effects of arachidonate-rich fungal oil and fish oil on murine hepatic and hippocampal gene expression. Lipids Health Dis 2002;1:2.
44 Sreekumar R, Unnikrishnan J, Fu A, Nygren J, Short KR, Schimke J, Barazzoni R, Nair KS: Impact of high-fat diet and antioxidant supplement on mitochondrial functions and gene transcripts in rat muscle. Am J Physiol 2002;282:E1055–E1061.
45 Welle S, Brooks AI, Delehanty JM, Needler N, Thornton CA: Gene expression profile of aging in human muscle. Physiol Genomics 2003;14:149–159.
46 Sreekumar R, Halvatsiotis P, Schimke JC, Nair KS: Gene expression profile in skeletal muscle of type 2 diabetes and the effect of insulin treatment. Diabetes 2002;51:1913–1920.
47 Yang YS, Song HD, Li RY, Zhou LB, Zhu ZD, Hu RM, Han ZG, Chen JL: The gene expression profiling of human visceral adipose tissue and its secretory functions. Biochem Biophys Res Commun 2003;300:839–846.
48 Nimgaonkar A, Sanoudou D, Butte AJ, Haslett JN, Kunkel LM, Beggs AH, Kohane IS: Reproducibility of gene expression across generations of Affymetrix microarrays. BMC. Bioinformatics 2003;4:27.
49 Kothapalli R, Yoder SJ, Mane S, Loughran TP Jr: Microarray results: how accurate are they? BMC. Bioinformatics 2002;3:22.
50 Gygi SP, Corthals GL, Zhang Y, Aebersold R: Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc Natl Acad Sci USA 2000;97:9390–9395.
51 Fields S, Song O: A novel genetic system to detect protein-protein interactions. Nature 1989;340:245–246.
52 Frederickson RM: Macromolecular matchmaking: advances in two-hybrid and related technologies. Curr Opin Biotechnol 1998;9:90–96.
53 Cagney G, Uetz P, Fields S: High-throughput screening for protein-protein interactions using two-hybrid assay. Methods Enzymol 2000;328:3–14.
54 Cagney G, Uetz P, Fields S: Two-hybrid analysis of the *Saccharomyces cerevisiae* 26S proteasome. Physiol Genomics 2001;7:27–34.
55 Fitzgerald RJ, Meisel H: Milk protein hydrolysates and bioactive peptides; in Fox PF, McSweeney PLH (eds): Advanced Dairy Chemistry. Amsterdam, Kluwer Academic/Plenum, 2003, pp 675–698.
56 Kaminogawa S, Totsuka M: Allergenicity of milk proteins; in Fox PF, McSweeney PLH (eds): Advanced Dairy Chemistry. Amsterdam, Kluwer Academic/Plenum, 2003, pp 647–674.
57 Malik Z, Bottomley R, Austen B: Allergenic properties of the genetic variants A and B of bovine beta-lactoglobulin. Int Arch Allergy Appl Immunol 1988;86:245–248.
58 Willige RWG, Fitzgerald RJ: Tryptic and chymotryptic hydrolysis of beta-lactoglobulin A, B, and AB at ambient and high pressure. Milchwissenschaft 1995;50:183–186.
59 Huang XL, Catignani GL, Swaisgood HE: Comparison of the size and rate of formation of peptides released by limited proteolysis of beta-lactoglobulins A and B with immobilized trypsin. J Agric Food Chem 1994;42:1281–1284.
60 Brazma A: On the importance of standardisation in life sciences. Bioinformatics 2001;17:113–114.
61 Desiere F, German B, Watzke H, Pfeifer A, Saguy S: Bioinformatics and data knowledge: the new frontiers for nutrition and foods. Trends Food Sci Technol 2002;12:215–229.
62 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, et al: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–29.
63 Ashburner M, Ball CA, Blake JA, Butler H, Cherry JM, Corradi J, Dolinski K, Eppig JT, Harris MA, et al: Creating the gene ontology resource: design and implementation. Genome Res 2004;11:1425–1433.
64 Kelso J, Visagie J, Theiler G, Christoffels A, Bardien S, Smedley D, Otgaar D, Greyling G, Jongeneel CV, et al: eVOC: A controlled vocabulary for unifying gene expression data. Genome Res 2003;13:1222–1230.
65 Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox, T, Cuff J, Curwen V, et al: The Ensembl genome database project. Nucl Acids Res 2002;30:38–41.