Biomarkers: background, classification and guidelines for applications in nutritional epidemiology

Dolores Corella\textsuperscript{1,2} and José M. Ordovás\textsuperscript{3,4,5}

\textsuperscript{1}Unidad de Epidemiología Genética y Molecular. Departamento de Medicina Preventiva y Salud Pública. Ciencias de la Alimentación, Toxicología y Medicina Legal. Universidad de Valencia. \textsuperscript{2}CIBER Fisiopatología de la Obesidad y Nutrición. Instituto de Salud Carlos III, Madrid. \textsuperscript{3}Department of Cardiovascular Epidemiology and Population Genetics, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid. \textsuperscript{4}IMDEA Alimentación, Madrid. Spain. \textsuperscript{5}Nutrition and Genomics Laboratory, JM-USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA. USA.

Abstract

One of the main problems in nutritional epidemiology is to assess food intake as well as nutrient/food component intake to a high level of validity and reliability. To help in this process, the need to have good biomarkers that more objectively allow us to evaluate the diet consumed in a more standardized, valid and precise way has often been commented upon. There are various definitions of biomarkers and also different classifications of the same. In general a biomarker can be defined as a characteristic that can objectively measure different biological samples and that can be evaluated as an exposure marker of normal or pathogenic biological processes or of responses to a certain intervention. The biological samples most commonly used in nutritional epidemiology are blood, red blood cells, plasma, serum, urine, nails, saliva, faeces and samples of different tissues. Exposure biomarkers (dietary intake), biomarkers of effects and biomarkers of disease status can be determined from these samples. In turn, exposure biomarkers can be temporarily categorized into markers of acute, medium term or chronic effects. Many difficulties arise in identifying good biomarkers. Currently, advances in omics are opening up new possibilities for obtaining new biomarkers of various kinds, using genomics, epigenomics, transcriptomics, lipidomics, proteomics and metabolomics. We shall review the present situation of biomarkers in nutritional epidemiology as well as the future trends of the new omic biomarkers.

(Nutr Hosp 2015;31(Supl. 3):177-188)

DOI:10.3305/nh.2015.31.sup3.8765

Key words: Biomarker, Diet, Intake, Biological sample, Genomics, Transcriptomics, Metabolomics.

Correspondence: Dolores Corella.
Departamento de Medicina Preventiva y Salud Pública.
Facultad de Medicina.
46010-Valencia.
E-mail: dolores.corella@uv.es

Biomarcadores: antecedentes, clasificación y guía para su aplicación en epidemiología nutricional

Resumen

En los estudios de epidemiología nutricional uno de los principales problemas es conocer la ingesta de alimentos y sus componentes de manera válida y precisa. Para ayudar en este proceso se ha planteado repetidas veces la necesidad de contar con buenos biomarcadores, que de manera más objetiva nos permitan conocer de manera más estandarizada, válida y precisa la dieta consumida. Existen varias definiciones de biomarcador y también distintas clasificaciones de los mismos. En general un biomarcador es una característica que puede medir objetivamente en distintas muestras biológicas y que puede evaluarse como indicador de exposiciones, de procesos biológicos normales o patogénicos o de respuesta a una intervención determinada. Las muestras biológicas más utilizadas en epidemiología nutricional son sangre total, eritrocitos, plasma, suero, orina, uñas, saliva, heces y muestras de distintos tejidos. En estas muestras se pueden determinar biomarcadores de exposición (ingestia dietética), biomarcadores de efectos y biomarcadores de estado de enfermedad. A su vez los biomarcadores de exposición pueden categorizarse temporalmente en biomarcadores de efectos agudos, a medio plazo y crónicos. Existen muchas dificultades en la identificación de buenos biomarcadores. Actualmente los avances en las nuevas omicás están abriendo nuevas posibilidades para la obtención de nuevos biomarcadores de distintos tipos utilizándose genómica, epigenómica, transcriptómica, lipidómica, proteómica y metabolómica. Revisaremos el estado actual de los biomarcadores en epidemiología nutricional así como las tendencias futuras de los nuevos biomarcadores omícicos.

(Nutr Hosp 2015;31(Supl. 3):177-188)

DOI:10.3305/nh.2015.31.sup3.8765

Palabras clave: Biomarcador, Dieta, Ingesta, Muestra biológica, Genómica, Transcriptómica, Metabolómica.
Background

The limitations of questionnaires are widely recognized when it comes to measuring dietary intake with sufficient validity and accuracy. Although efforts are made to improve the validity of those tools by using dietary records or various 24-h recalls instead of the less accurate food frequency questionnaires, random and systematic bias, which make self-reported dietary measurements diverge from the reality, always exists. Moreover, these errors in measuring food intake extend to nutrients and other food components derived from the food consumed, not only because the intake has not been noted down with sufficient validity and accuracy, but because there are other factors such as variability in the composition of the food consumed, etc. which also contributes to nutrients and food components derived theoretically from food composition tables not being faithfully adjusted to the real consumption. Measuring these intakes well is very important because, on most occasions, nutritional studies not only have the aim of getting to know the dietary intake of a certain population, but also a second step which is to study the associations between food intake and a certain health problem. It is, therefore, clearly true that the accurate assessment of dietary exposure is crucial in investigating associations between diet and disease. Hence, other alternative means are required for getting to know the food and nutrient intake (and non-nutritive food components) contributed by the same with greater validity and accuracy than that obtained through self-reported measures.

Nutritional biomarkers are important for future research into associations between diet and health, as they can provide an objective assessment method for dietary exposure. However, the definition of biomarker is not simple and there are indeed many definitions depending on the application of those biomarkers. A widely used definition of biomarker was provided by the Biomarker Definition Working Group (BDWG) in 2001. According to this, a biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic intervention. However, this definition of biomarker is not well suited to all situations and many variations of the same have been proposed. In nutritional research, we have to use a broad definition and also one that is adaptable to each situation as we will need biomarkers that cover at least the following aspects: dietary intake, nutritional status, nutrient exposure, effects of nutritional interventions on physiological and/or pathological outcomes and to provide information on inter-individual differences in response to diet. We should also bear in mind that many biomarkers can fall into more than one of these categories.

Currently, the study of nutritional biomarkers, whether they be biochemical, functional, or clinical indices of nutrient intake or of metabolism, are revolutionizing our understanding of the role of nutrients and food components in health and disease. Although there is huge interest in the use and development of new biomarkers, the present situation is that we still do not have good biomarkers for most of the previously mentioned important aspects. This is so much so that various public and private research organizations are stressing the need to delve deeper into the investigation of new nutritional biomarkers and are providing an impetus to research along these lines. Thus, since one of the initial conclusions of the European Commission-funded project PASSCLAIM, coordinated by ILSI Europe, was that there is a need for adequate markers in nutrition sciences, research into better biomarkers (among them the recent initiatives of the Joint Programming Initiative (JPI) on food and nutrition in Europe, which this year began financing two new projects (MIRDIE and FOODBALL), has been considered a priority. These projects aim at the validation of biomarkers and the investigation of intake/exposure and nutritional status of biomarkers in the area of nutrition and health, the former being focused on microRNAs and the latter on the application of metabolomics. The United States is now promoting projects aimed at research into nutritional biomarkers and in April, 2012, the Sackler Institute for Nutrition Science and the New York Academy of Sciences organized a conference entitled Biomarkers in Nutrition: New Frontiers in Research and Application. The aim of this conference was to get scientists and practitioners from industry, academia, and government organizations to work together on assessing the current state of knowledge about nutritional biomarkers, to identify important challenges and unanswered questions, and to catalyze new research in the field so that soon it may be possible to implement good biomarkers in nutritional epidemiology that will enable a better measuring of food intake, its effects, and its association with states of health-disease.

In this article, we shall review present knowledge on biomarkers in nutritional epidemiology and will go deeper into the new omics as these are poised to revolutionize the identification of new markers in nutritional studies.

Classification of biomarkers and guidelines for use

The use of biomarkers was first described by Isaakson in 1980 when he proposed urinary nitrogen as an independent measure of protein intake and it remains one of the most common biomarkers used. However, not all biomarkers share the same characteristics. There are different classifications of biomarkers. Potischman defined a biomarker as “any biological specimen that is an indicator of nutritional status with respect to intake or metabolism of dietary constituents. It can be biochemical, functional or clinical index of status of an essential nutrient or another dietary constituent”. This author classified biomarkers into two large groups:
Biomarkers of nutritional exposure and biomarkers of nutritional status. Exposure biomarkers would be those used for validating dietary measurement, or as a surrogate of dietary intake. Both those and the biomarkers of nutritional status should be evaluated according to precision, accuracy, sensitivity, specificity to the nutrient, variability between subjects and temporality. However, this definition can be extended on taking into account that in nutritional studies, we are not only interested in measuring the diet well, but also its relationship with states of health-disease, so, in nutritional studies, it is also necessary to incorporate the measurements of biomarkers related with disease so as to have more complete determinations. Table I shows this extended classification of biomarkers in nutritional studies for these purposes. The nutritional biomarkers’ role in health and disease has evolved from markers of deficiency in one specific disease (e.g. vitamin A and eyes), to a multitude of chronic conditions spanning the endocrine, cardiovascular, respiratory, digestive, immune, and nervous systems, among others.

However, biomarkers can also be classified depending on their temporality. Thus, biomarkers can be categorized into short-term (reflecting intake over past hours/days), medium-term (reflecting intake over weeks/months) and long-term biomarkers (reflecting intake over months/years). The type of biological sample used for the analysis of these biomarkers is a main determinant. For example, biomarkers measured in urine, plasma or serum reflect short-term intake well, whereas measurements of biomarkers in red blood cells or in adipose tissue are markers of medium-term intake. Likewise, biomarker measurements in hair, nails, or teeth, are more often employed as long-term biomarkers.

Another classification of biomarkers distinguishes between recovery, concentration, replacement and predictive biomarkers:

**Recovery biomarkers** are based on the concept of the metabolic balance between intake and excretion over a fixed period of time and then provide an estimate of absolute intake levels. Recovery biomarkers are specific biologic products that are directly related to intake and not subject to homeostasis or substantial inter-individual differences in metabolism. Only a few recovery biomarkers are known. The best examples of recovery biomarkers are as follows: doubly labeled water which is utilized to measure the metabolic rate and total energy expenditure; urinary total nitrogen/potassium which are utilized to estimate total daily protein consumption and potassium intake, respectively. The first large validation study with recovery biomarkers was the Observing Protein and Energy Nutrition (OPEN) Study, conducted by the National Cancer Institute in 1999-2000.

Among the other later studies into these recovery markers, we should mention the work of Dr. Prentice and his group on a sub-sample of women who were participating in the Women’s Health Initiative Dietary Modification Trial (WHI-DM). The WHI-DM is a randomized controlled trial among postmenopausal women aimed at analyzing whether a low-fat diet reduced the incidence of breast and colorectal cancer, and secondarily, heart disease. The subset of women, completed a food frequency questionnaire (FFQ), the doubly labeled water protocol, and a 24-hour urine collection (as biomarker for protein consumption). The collection of these recovery biomarkers allowed the researchers to characterize the measurement error distributions of energy and protein assessed by the FFQs. In addition, they were able to identify the general characteristics (age, sex, obesity, etc) of the participants who made the measurement greater errors in the FFQs. In this specific study, they were able confirm that FFQs underreported the consumption of energy and protein. They also reported that the energy underreporting was greater among overweight/obese women and younger women. These results enabled them to create regression calibration equations for energy and protein and to apply them to the measurements obtained from the FFQs.

**Concentration biomarkers** are biomarkers that have a correlation with intake, but because they are affected by metabolism or personal characteristics (sex, age, smoking, obesity, etc), they cannot be used as measures of absolute intake or for assessing error

| Table I |
| Classification of biomarkers in nutritional studies |
| Biomarkers of dietary exposure |
| Different types of biomarkers aimed at assessing dietary intake of different foods, nutrients, non-nutritive components or dietary patterns (recovery biomarkers, concentration biomarkers, recovery biomarkers and predictive biomarkers). Example: Urinary nitrogen as biomarker of protein intake. |
| Biomarkers of nutritional status |
| Biomarkers which reflect not only intake but also metabolism of the nutrient(s) and possibly effects from disease processes. Example: Some of the biomarkers of one-carbon metabolism such as homocysteine, which reflect not only nutritional intake, but also metabolic processes. It is important to note that a single biomarker may not reflect the nutritional status of a single nutrient, but may indicate the interactions of several nutrients. |
| Biomarkers of health/disease |
| Biomarkers related to different intermediate phenotypes of a disease or even to the severity of the disease. Example: plasma concentrations of total cholesterol or triglycerides associated for cardiovascular diseases. |
of self-reported intakes in validation studies\textsuperscript{12}. Examples of concentrations biomarkers are as follows: Serum carotenoids, lipids, vitamins, etc. They can be used to analyze the relationship between the concentration of the above in a certain tissue and variables if health states\textsuperscript{6}.

Replacement biomarkers are closely related to concentration biomarkers and often the distinction between them is difficult to make. Their differentiating characteristic is that they refer specifically to compounds for which information in food composition databases is unsatisfactory or unavailable. Examples of these replacement biomarkers are some aflatoxins, some phytoestrogens\textsuperscript{13}, or some of the recent biomarkers identified through metabolomics\textsuperscript{6}, which we shall refer to later on in this review.

Recently, a newer classification of biomarkers, termed predictive biomarkers, has been proposed. These biomarkers show a dose-response relationship with intakes. Like recovery biomarkers, predictive biomarkers are sensitive, time dependent, show a dose-response relationship with intake levels and may be affected by personal characteristics but the difference is that their overall recovery is lower\textsuperscript{14}. Examples of predictive biomarkers are 24-hour urinary fructose and sucrose\textsuperscript{14}. It has also been reported that the use of these Urinary Sugars Biomarker is useful for assessing measurement error in self-reported sugars Intake in the Nutrition and Physical Activity Assessment Study (NPAAS)\textsuperscript{15}. Figure 1 summarizes the classification of the above biomarkers as well as their applications in validating dietary assessment methods, measurement error and estimating associations with the different phenotypes of disease.

Biomarkers require the obtaining of different types of biological samples for their measurements. The most commonly employed are blood, urine and saliva, although increasing more determinations are being taken from other sample such as faeces, hair, nails, adipose tissue and other specific tissues depending on the aims of the study. In the following section, some general comments are made on obtaining and storing biological samples for determining biomarkers in epidemiological studies.

Obtaining and storing biological samples in nutritional epidemiology studies

Given that nowadays several biomarkers are available for use in nutritional epidemiology studies and, in the future, it is expected that many more biomarkers may be incorporated, it is advisable to obtain and store biological samples in any new nutritional epidemiology studies that are initiated. The number of biological samples and their complexity will depend on the aims of the epidemiological studies and the means available for them. It is advisable to at least take biological samples of the saliva and urine of all the participants in a study, as these are the least invasive of biological samples. From these, a considerable number of biomarkers can be determined and from saliva it is even possible to isolate DNA from the bucal cells that are gathered in the same. It is also advisable to not take a single biological sample in the same tube, but to divide it into different aliquots in order to prevent the processes of freezing and defreezing that can affect a number of the biomarkers that one wishes to determine. A good strategy, therefore would be to store a minimum of two or three aliquots for each participant. These samples would have to be frozen at a very low temperature (deep freezing at -80ºC is the most common conservation) in order to ensure the better conservation and avoid the degradation of the biomarkers. If it is not possible to obtain biological samples from all participants, it would at least be advisable to obtain them from a representative sub-sample of the population. If it is possible to obtain and store more samples, it is recommendable to undertake extractions of fasting peripheral venous blood and then process them by means of centrifugation, etc. to obtain serum, plasma and buffy-coat aliquots. These samples will be very valuable later on for determining different biomarkers. Although freezing at -80ºC may be sufficient, it would be ideal to freeze the samples at a lower temperature and store them in liquid nitrogen\textsuperscript{15}. However, this type of conservation is not widely available and is limited to a few studies). If, for the research in question, the measurement of biomarkers in red blood cells or in other types of blood cells such as leucocytes is considered to be important, these samples have to be isolated from the extracted blood through the standard protocols and frozen separately from the other components. In the same way, if it is decided to undertake determinations with the DNA or RNA of the participants, these samples will also have to be obtained in a standardized way through the pertinent protocols. Temporality in sample collecting will depend on whether the study is cross-sectional or longitudinal. In cross-sectional studies it will only be necessary to gather them once, whereas in longitudinal studies, samples have to be taken at baseline and at different moments of follow-up depending on the aims of the study. The number of aliquots obtained may be very large in studies that include thousands of participants, so there a good labelling and traceability infrastructure of the frozen stored samples will have to be planned. It is also advisable to think about building some kind of biobanks and to follow the protocols for such purposes\textsuperscript{15}. On choosing the protocols for obtaining and storing samples, one must bear in mind several limitations for the subsequent validity of the determinations and the comparability of results. Currently, there are different anti-coagulants that are used in blood sample collection tubes. Later determinations may vary depending on whether citrate, heparin or EDTA are used. For omic studies, one of the limiting samples is obtaining DNA, as this requires prior isolation in fresh samples or blood samples to be collected and stored in the pre-
sence of RNA preservatives. Although obtaining DNA for genotyping does not present any problems and generally fulfils the quality requirements regardless of the anticoagulant used, storage time, etc., the obtaining of DNA for epigenetic studies such as methylation can be subject to more problems of validity and reproducibility depending on the method employed, the time of the year in which the sample was collected, etc. For further details of the factors that affect the conservation and processing of biological sample in omic studies, the work of Hebels et al.\(^1\) is recommended, in which some general guidelines are provided in the context of the European project EnviroGenomarkers (http://www.envirogenomarkers.net). In this Project, blood-derived biobank samples are being analyzed on multiple omic platforms with the aim of discovering new biomarkers of exposure and disease risk. Furthermore, in addition to the good storage of biological samples, valid laboratory procedures for analyzing different biomarkers are required that allow for comparisons among laboratories. Without such comparability, it is not possible to make recommendations regarding appropriate levels.

Use of dietary biomarkers in combination with data from questionnaires

Another approach that can be used is the joint use of intake biomarkers with data coming from questionnaires where self-reported intake is measured. This combination allows one to increase the validity of measurements and also increase the statistical power of the subsequent diet-disease associations. This approach has been used by various authors, among them Freedman et al.\(^1\) in the CAREDS study (Age-Related Eye Disease Study). It is an ancillary study of the Women’s Health Initiative (WHI) Observational Study, a prospective cohort study of 93,676 postmenopausal women, recruited from 40 sites around the United States. Specifically, the researchers used this study to illustrate that the inverse association between dietary lutein plus zeaxanthin and nuclear cataracts, improved on using biomarkers in an overall way for these carotenoids with data intake from self-reported questionnaires. Dietary intake was assessed by using the WHI semi-quantitative FFQ. Serum samples were collected from a random and systematic dietary measurement errors addressed by:

- Improvements in diet assessment methods
- Refinement of existing methods
- Enhancement of food composition databases
- Innovation of new methods for biomarker integration
- Integration of dietary and biomarker data for identification and correction of measurement errors.

### Diagram

**Fig. 1.—Classification of biomarkers and their applications in validating dietary assessment methods, and estimating associations with the different phenotypes of disease (adapted from reference\(^6\)).**

<table>
<thead>
<tr>
<th>Dietary Assessment</th>
<th>Recovery Dietary Biomarkers</th>
<th>Improvements for Dietary Biomarkers in Current Use:</th>
</tr>
</thead>
</table>
| Estimation of dietary intakes (e.g. Food frequency questionnaires, 24-hour recalls, etc) | Excretion levels are highly correlated with intake. Gold standard. I.E. Doubly labeled water, urinary nitrogen or potassium. | - Enhanced laboratory methods  
- Better understanding of:  
  - Metabolic pathways  
  - Genome-diet interactions  
  - Epigenome-diet interactions  
  - Gene-Gene interactions |

| Concentration Dietary Biomarkers | Good correlation with intake but lower than for recovery biomarkers. e.g. Serum vitamins, carotenoids, etc. |

| Replacement Dietary Biomarkers | Serve as a proxy for intake when it is not possible to capture intake due to limited nutrient databases to assess intake. e.g. phytoestrogens, polyphenols or aflatoxin |

| Predictive Dietary Biomarkers | Show a dose-response relationship with intake levels but the distinction is that their overall recovery is lower e.g. urinary sucrose or fructose |

| Identification of New Dietary Biomarkers based on “omics”: | Nutritional Metabolomics (lipidomics or proteomics) Identification of metabolic profiles as biomarkers specific to different dietary intake  
Nutritional Epigenomics and Transcriptomics |

| Effect on Disease Risk | On intermediate and/or disease phenotypes (or Disease biomarkers) |

| Enhanced Application and Interpretation of Dietary Biomarkers |

| Incorporation of individual polymorphisms, whole genome, transcriptomics, epigenomics, Metabolomics and other omics |

Biomarkers: background, classification and guidelines for applications in nutritional epidemiology 181
after 10 or more hours of fasting at the WHI baseline examinations and were analyzed for lutein and zeaxanthin (sum of their trans isomers). The authors investigated 3 ways of analyzing reported dietary lutein plus zeaxanthin intake and lutein plus zeaxanthin serum level; the third one combining the self-report and biomarker measurements. For the third method, they used a combined score where data from both determinations were used. For further details of that score, the original reference should be consulted19. The conclusion to which the authors arrived is that by combining a biomarker of dietary intake with self-reported dietary intake can increase the statistical power for detecting a diet-disease association. Therefore, it is advisable to use this combined method whenever possible. However, they also recognize limitations to that combination when the biomarker is not valid or when the error derived from the food or nutrient measurement through the questionnaires also has great errors.

**Limitations and considerations in using biomarkers**

Although biomarkers can provide a more objective measurement of dietary intake, for many of them there are various inter-individual factors that could skew biomarker measures of dietary intake and give untrue values20. Among these factors, apart from sex, age, tobacco smoking, alcohol consumption, drugs, physical activity and other lifestyle factors, are other factors of diet (nutrient-nutrient interactions), the type of biological sample (e.g. blood, plasma, serum, urine, etc.) and conditions related with the obtaining and storage of the samples (conditions of sample collection, transport, treatment, storage conditions, length of storage, time of collecting the samples including the day of the week, season of the year, etc.), as well as the particularities of the laboratory methodology for determining them (precision, accuracy, detection limits of the analytical technique and inter-laboratory variations)20. Apart from all these, genetic markers are becoming increasingly important as research advances on inter-individual variability, in particular, genetic polymorphisms in relevant genes related with each one of the biomarkers analyzed. There are many types of genetic polymorphisms. The most commonly studied are those that consist of a single base change in a genome site and these are called single nucleotide polymorphisms (SNPs). There are millions of SNPs in the human genome and the technology for their determination has developed spectacularly, moving on from rudimentary techniques that were slow and expensive to other much more automated technologies at a lower cost. It is, therefore, now very easy and fast to incorporate genetic determinations into nutritional epidemiology studies6. This genetic variation may not only affect the preferences in the choice and consumption of food, but may also play an important role in nutrient metabolism and in bio-availability, absorption, transport, bio-transformation, and excretion of nutrients or food components. There are many examples of the influence of genetic polymorphisms in the concentrations of different biomarkers. Among them we should mention the influence of the genetic polymorphism rs1279683 (A>G) in the SLC23A2 gene and plasma concentrations of vitaminC21. Plasma concentrations of vitamin C are determined by dietary intake, as well as by genetic factors. L-ascorbic acid obtained from the diet is transported across the cell membrane by sodium L-ascorbic acid cotransporters (SVCTs). Two isoforms, SVCT1 (encoded by the SLC23A1 gene) and SVCT2 (encoded by the SLC23A2 gene), play central roles in the absorption and accumulation of vitamin C in many tissues. In a study undertaken by our group to investigate the influence of vitamin C plasma concentrations on the risk of glaucoma21, we found that the rs1279683-SLC23A2 SNP was strongly associated with plasma vitamin C concentrations both in cases and controls (Fig. 2). According to these results, homozygous carriers of the variant G allele have significantly lower plasma vitamin C concentrations than the other genotypes despite having similar intakes. Another relevant example of the influence of genetic polymorphisms on concentrations of a biomarker regardless of intake are the polymorphisms in relevant genes in the metabolism of polyunsaturated fatty acids22. The delta-6 desaturase (D6D) and delta-5 desaturase (D5D) are membrane-bound enzymes that catalyze the rate-limiting formation of long-chain polyunsaturated fatty acids. The desaturase-encoding genes (FADS1 for D5D and FADS2 for D6D) form a gene cluster on chromosome 11 together with a third desaturase gene, FADS3, of lesser known function. Several studies have consistently replicated the associations between polymorphisms in the FADS1 and FADS2 genes and polyunsaturated fatty acid concentration measurements in different biological samples22. This relevant genetic influence is important to bear in mind in epidemiological studied and, as a guideline, it is recommended to gradually incorporate determinations of the most relevant genetic polymorphisms in nutritional epidemiology studies as a control of the most important inter-individual differences.

**New omic-based biomarkers**

Whether as control for inter-individual differences in measuring classic biomarkers or properly considered as biomarkers, omic technologies have led to the study and validation of new biomarkers in nutrition and health23. Among them are those shown in table II:
- **Genetic biomarkers:** These biomarkers are based on the determination of genetic polymorphisms (mainly SNPs) and can be either of intake or of effect (metabolism) or as disease risk. They can be determined in the DNA of any biological sample that contains cells with a nucleus (advantage). Their determination does not vary over time and the samples are easily conserved and transported (blood, urine, hair, different tissues, etc. can all be used). In addition, they allow a rapid determination at low economic cost. Furthermore, in recent years high density genotyping arrays have been available that have allowed us to simultaneously determine thousands of genetic polymorphisms. This ability has led to the so-called genome-wide association studies (GWAs) and to the discovery of new genes and SNPs associated with the different levels of the other biomarkers, dietary intake of disease phenotypes. Recently a number of meta-analyses including thousands of individuals have been published identifying new gene variants associated with food and nutrient intake\(^{24,25}\) or with biomarker concentrations, such as circulating phylloquinone\(^{26}\). Also recently, genomic studies into lipids have incorporated the technologies of next generation sequencing for identifying new genetic variants associated with different biomarkers, mainly identifying new low prevalence variants\(^{27}\).

In the other hand, Mendelian randomization involving genetic biomarkers is currently used as a technique for assessing causal associations in observational data. Genetic variants associated with the risk factor of interest are regarded in a similar way to random assignment in a clinical trial. The modern meaning of Mendelian randomization is based on Mendel’s second law, the law of independent assortment, which assumes that the inheritance of one trait is independent of the inheritance of other traits\(^{28}\). In terms of nutritional bio-

---

**Table II**

<table>
<thead>
<tr>
<th>Classification of new omic-based biomarkers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic biomarkers</td>
</tr>
<tr>
<td>Epigenetic biomarkers</td>
</tr>
<tr>
<td>Transcriptomic biomarkers</td>
</tr>
<tr>
<td>Proteomic biomarkers</td>
</tr>
<tr>
<td>Lipidomic biomarkers</td>
</tr>
<tr>
<td>Metabolomic biomarkers</td>
</tr>
</tbody>
</table>
markers, SNPs that have a well-characterized biological function can be utilized to study the effect of a suspected environmental exposure on disease risk. Thus, there are several studies using genetic variants as proxies (instrumental variables) for environmental exposures. A well known example is the lactose polymorphism. The The -13910C>T polymorphism (rs4988235) upstream from the lactase (LCT) gene is strongly associated with lactase persistence (LP) in Europeans. Lactate nonpersistent (LNP) individuals have difficulty in metabolizing lactose and, after consuming dairy products, often have symptoms of abdominal pain and diarrhea. As a result, individuals with LNP tend to consume less lactose-containing dairy products and, therefore, the variant associated with LNP (CC genotype) can be a proxy for low exposure to milk. Likewise, another informative gene is the ALDH2 (aldehyde dehydrogenase). Acetaldehyde is the first metabolite of ethanol. ALDH2 is the enzyme primarily responsible for the elimination of acetaldehyde. There is a functional polymorphism (Glu487Lys), resulting in an inactive enzyme. The variant 487Lys allele is associated with reduced ability to metabolize acetaldehyde, and thus contributes to high acetaldehyde concentrations after alcohol drinking (resulting in facial flushing, nausea and headache in response to consumption of alcohol). Because of these adverse reactions, subjects homozygous for the 487Lys allele drink noticeably less alcohol than homozygous subjects for the wide-type allele. Thus, ALDH2 polymorphism can define groups with different amounts of alcohol intake and these variants used as a surrogate of alcohol exposure. However, although the Mendelian randomization approach shows considerable promise in integrating genetic markers into nutritional epidemiology research, its application to other genetic variants has several potential limitations when the Mendelian randomization assumptions are violated. In addition to these considerations, genetic biomarkers are crucial in determining intermediate (plasma lipids, fasting glucose, oxidative markers, inflammation markers, etc) and incidence of disease (cardiovascular diseases, cancer, neurodegenerative diseases, type 2 diabetes, obesity, etc). In nutritional epidemiology, when establishing the association between diet and diseases, the most relevant genetic polymorphisms related with the phenotypes of interest should be determined. Currently there are hundreds of SNPs consistently associated with different phenotypes of nutrition-related diseases that should be considered in studies focused on these phenotypes. By way of example table III shows the main genes and genetic variants consistently associated with intermediate and disease phenotypes in cardiovascular epidemiology. The integration of the genetics into nutrition has propelled the advances in gene-diet interactions and Nutritional Genomics. In the PREDIMED study we have found some interesting gene-diet interactions in determining both intermediate and cardiovascular disease phenotypes involving common polymorphisms in the TCF7L2 and the MLXIPL genes and intervention with Mediterranean diet.

- **Epigenetic biomarkers:** The term epigenetics/epigenomics is used to describe a variety of modifications to the genome that do not involve changes in the DNA sequence and can result in alteration of gene expression allowing for differential expression of common genetic information. It constitutes the missing link among genetics, the environment and disease. One of the main advantages of the epigenetic biomarkers, unlike the variations in the genome, is that the epigenetic markers are reversible and may allow a rapid adaptation to the environment. There are 3 main categories of epigenetic marks: DNA methylation, histone modification and noncoding RNA. DNA methylation is a well characterized epigenetic modification of the genome. Most DNA methylation in humans occurs at cytosine–phosphate–guanine (CpG) dinucleotides and consists of the addition of a methyl group on position 5 of cytosine residues of the CpG island, providing marks in the genome by which genes are set to be transcriptionally activated or silenced. Hypermethylation or hypomethylation of relevant islands have been associated with several phenotypes of disease. Preliminary studies exist that show that the diet can affect the methylation of certain sites of the DNA and that these changes in methylation are dynamic. However, many more studies are required to establish them as new biomarkers of intake or disease. As regards epigenetic regulation by non-coding RNAs and although different classes of non-coding RNAs such as long non-coding RNAs, small nucleolar RNAs ( snoRNAs) microRNA (miRNAs) have been characterized, miRNAs are currently the most important. miRNAs are small (18-25 nucleotide) functional non-coding RNAs, that regulate gene expression of their target mRNA in a post-transcriptional manner and have emerged as crucial epigenetic regulators of many process related to nutrition. Recent evidence highlights how diet may influence several disease phenotypes through modulation of miRNA expression. Additionally, circulating miRNAs are emerging as biomarkers of several diseases. Moreover, There are a few studies that indicate that some exogenous microRNAs could be used as biomarkers of food intake (detection of miRNAs of rice consumed in human plasma), but the-
### Table III

<table>
<thead>
<tr>
<th>References</th>
<th>Gene</th>
<th>Genetic variant</th>
<th>Intermediate phenotype</th>
<th>Cardiovascular disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khan et al, 2013</td>
<td>APOE</td>
<td>C2, C3 and C4 polymorphism (rs4420638 and rs7412)</td>
<td>Higher LDL-C and total cholesterol in APOE-E4 allele carriers in comparison with E3/E3 subjects</td>
<td>The APOE-E4 allele associated with higher risk of cardiovascular diseases (stroke and myocardial infarction)</td>
</tr>
<tr>
<td>Ridker et al, 2009</td>
<td>CETP</td>
<td>Several common SNPs in partial LD: rs708272, rs7202364 and rs4329913</td>
<td>Higher HDL-C and APOA1 concentrations in carriers of the minor allele</td>
<td>Lower myocardial infarction risk in carriers of the minor allele</td>
</tr>
<tr>
<td>Wang et al, 2011</td>
<td>LPL</td>
<td>Several common SNPs in partial LD: rs328 and rs230</td>
<td>Higher HDL-C and lower triglyceride concentrations in carriers of the minor allele</td>
<td>Lower stroke risk in carriers of the minor allele</td>
</tr>
<tr>
<td>Zang et al, 2011</td>
<td>APOA5</td>
<td>Common SNP: -1131T&gt;C (rs662799) and S19W (rs3135506)</td>
<td>Higher triglyceride concentrations in carriers of the minor alleles</td>
<td>Higher risk of coronary artery disease in carriers of the minor alleles</td>
</tr>
<tr>
<td>Teslovich et al, 2010</td>
<td>LDL-R</td>
<td>SNP rs6511720</td>
<td>Lower LDL-C and total cholesterol in carriers of the minor allele</td>
<td>Lower risk of coronary artery disease in carriers of the minor allele</td>
</tr>
<tr>
<td>Teslovich et al, 2010</td>
<td>CILP2</td>
<td>SNP rs10401969</td>
<td>Lower LDL-C and total cholesterol in carriers of the variant allele</td>
<td>Lower risk of coronary artery disease in carriers of the variant allele</td>
</tr>
<tr>
<td>Teslovich et al, 2010</td>
<td>SORT1</td>
<td>SNP rs629301</td>
<td>Lower LDL-C and total cholesterol in carriers of the minor allele</td>
<td>Lower risk of coronary artery disease in carriers of the minor allele</td>
</tr>
<tr>
<td>Teslovich et al, 2010</td>
<td>KLF14</td>
<td>SNP rs4731702</td>
<td>Higher HDL-C in carriers of the variant allele</td>
<td>Lower risk of coronary artery disease in carriers of the variant allele</td>
</tr>
<tr>
<td>Kathiresan and Srivastava, 2012</td>
<td>TRIB1</td>
<td>Several common SNPs in partial LD: rs2954029, rs2954022 and rs2980885</td>
<td>Variant allele is associated with lower triglycerides, lower LDL-C cholesterol, and higher HDL-C</td>
<td>Carriers of the variant allele have lower risk of coronary heart disease</td>
</tr>
<tr>
<td>Kathiresan and Srivastava, 2012</td>
<td>PCSK9</td>
<td>A common missense variant is associated with lower LDL-C</td>
<td>Variant allele is associated with lower LDL-C</td>
<td>Carriers of the variant allele have lower risk of coronary heart disease</td>
</tr>
<tr>
<td>Do et al, 2013</td>
<td>APOAI</td>
<td>SNP rs10790162</td>
<td>Variant allele is associated with higher triglycerides and LDL-C</td>
<td>Carriers of the variant allele have higher risk of coronary artery disease</td>
</tr>
<tr>
<td>Do et al, 2013</td>
<td>APOB</td>
<td>SNP rs1367117</td>
<td>Variant allele is associated with higher triglycerides and LDL-C</td>
<td>Carriers of the variant allele have higher risk of coronary artery disease</td>
</tr>
</tbody>
</table>
Transcriptomic biomarkers: Transcriptomics provides us with knowledge of the transcriptome, either individually for each specific gene studied or analyzing the expression of various genes simultaneously on different scales. In this way we can investigate how exposure to different dietary factors affects the expression of all genes (genome-wide transcriptome) or the specific genes. These studies of expression can be conducted either by analyzing intervention with whole diets (for example the Mediterranean diet as against a low fat diet) or by administering specific foods (e.g. olive oil) or specific components of the diet (vitamins, etc.). Although initially these transcriptomic studies were carried out independently of other omics, in recent years, the general trend is to integrate them with other omics: genomics, lipidomics/metabolomics and epigenomics. However, it must be borne in mind that one of the limitations of transcriptomics or epigenomics biomarkers is that the transcriptome and the epigenome are not the same for all the cells of the organism, as is also true for genomics, but that the level of expression varies depending on the tissues analyzed, adding a little more difficulty to the investigation into these biomarkers.

Proteomic, lipidomic and metabolomics biomarkers. Proteomics, lipidomics and metabolomics through the comprehensive study of proteins, lipids and metabolites are also beginning to be applied in the nutritional biomarker field, providing promising results. Metabolomics can be defined as the screening of small-molecule metabolites present in samples of biological origins. The characterization of all the metabolites can provide a picture of the metabolism and a molecular fingerprint. Such a characterization is a biomarker of a biological state of the subject. In addition, metabolomics can be used to examine the outcome of nutritional intervention strategies by observing and comparing metabolic marks. This science is still in its infancy but promises to revolutionize nutritional biomarkers. Until just recently, the analysis of food was limited to estimate its nutritional value based on the content: carbohydrates, fats, proteins, water, vitamins, and minerals. In addition, several non-nutritive components have been determining. However, metabolomics is helping to explore the thousands of additional components. A large proportion of the food metabolome consists of phytochemicals. Moreover, metabolomics is able to find a large list of environmental chemicals such as plaguicides and different toxins in foods and beverages. The detection of these compounds in nutritional studies may help to have a more holistic analysis of the influence of food on health and disease. There are several metabolomic studies showing promising results in the nutritional field. Guertin et al. carried out a metabolomic investigation in a relatively large (n=502) sample of participants in the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, with the aims of identifying metabolites that are biomarkers of usual dietary intake and to evaluated metabolite reproducibility and required sample sizes to ascertain the potential for metabolomics in nutritional epidemiology. Baseline serum was analyzed by using ultra-high-performance liquid-phase chromatography–mass spectrometry and gas chromatography–mass spectrometry. They detected 412 known metabolites. Green vegetables, citrus, red meat, shellfish, fish, peanuts, rice, butter, coffee, beer, liquor, total ethanol, and multivitamins were each correlated with at least one metabolite and in total, 39 dietary biomarkers were identified (see Tables in reference for more detail). They found some strong associations that replicated previous findings such as the correlation between citrus intake with stachydrine; coffee intake with trigonelline-N-methyl nicotinate and quinate) and alcohol intake with ethyl glucoronide. In the PREDIMED Study we also have used metabolomics for detecting new biomarkers of intake. In one of the metabolomic studies we assessed the effect of the Mediterranean diet intervention (supplemented either with extra-virgin olive oil or nuts) intervention in a sub-sample of non-diabetic participants. The 1H NMR urinary profiles were examined at baseline and after 1 and 3 years of follow-up. In comparison with the control group (low fat diet), the most prominent results for the Mediterranean diet groups was the identification of a characteristic profile related to the metabolism of carbohydrates (3-hydroxybutyrate, citrate, and cis-aconitate), creatine, creatinine, amino acids (proline, N-acetylglutamine, glycine, branched-chain amino acids, and derived metabolites), lipids (oleic and suberic acids), and microbial metabolites (phenylacetylglutamine and p-cresol). These results showed that the application of NMR-based metabolomics make possible the classification of individuals regarding their dietary pattern and the response to specific dietary interventions.

Overall, proteomics, lipidomics and metabolomics are in general being considered as the great innovation in the discovery of new biomarkers of intake, effect and pathology. Many advances are being made in this field and there are important findings. Although the results are generally speaking still preliminary and te-
techniques are still expensive for large epidemiological studies, these are considered to be important technologies of the future for detecting the intake of specific foods and other relevant biomarkers of health status.

Conclusions

Given the huge advances that methodologically are taking place in the field of biomarkers in nutritional studies with the incorporation of the new omics, it is advisable to obtain and store biological samples in nutritional epidemiology studies in order to be able to undertake determinations of the main biomarkers related with the aims of the study. Currently, there are still limitations to the validity and reliability of many markers, but in future it hoped to minimize many of these limitations and to have low cost validated biomarkers available.

Acknowledgments

This study has been supported by The Spanish Ministry of Health (Instituto de Salud Carlos III) and the Ministry of Economy and Innovation, Spain and Fondo Europeo de Desarrollo Regional (projects CIBER 06/03, CNIC-06, P11/02505, P11000462, AGL2010-22319-C03-03) and by the Generalitat Valenciana, Spain (ACOMP/2013/165 and ACOMP/2013/159).

Disclosures

The authors have no conflict of interests

References


Biomarkers: background, classification and guidelines for applications in nutritional epidemiology 187


