FOBI: An ontology to represent food intake data and associate it with metabolomic data

Pol Castellano-Escuder[1,2,3], Raúl González-Domínguez[1,3], David S. Wishart[4], Cristina Andrés-Lacueva[1,3] and Alex Sánchez-Pla[2,3]

[1]Biomarkers and Nutritional & Food Metabolomics Research Group, Department of Nutrition, Food Science and Gastronomy. University of Barcelona, Barcelona, Spain.
[2]Statistics and Bioinformatics Research Group, Department of Genetics, Microbiology and Statistics. University of Barcelona, Barcelona, Spain.

[3]CIBERFES, Instituto de Salud Carlos III. Madrid, Spain.
[4]Department of Biological Sciences, University of Alberta, Edmonton, AB, T6G 2E8, Canada.

Abstract

Nutritional research largely relies on accurate dietary assessment, which is of great relevance to evaluate food intake and dietary habits. Dietary assessments also help in understanding the association between nutrition and health status. Nutritional research is often conducted by using two complementary approaches: 1) self-reporting methods (e.g. food frequency questionnaires, dietary recalls) (1), and 2) the measurement of dietary biomarkers using a variety of analytical chemistry techniques, including metabolomics (2, 3). With regard to traditional dietary assessment tools, it should be noted that subjective self-reports generate very complex textual data, containing types and quantities of foods and recipes in very diverse and heterogeneous formats that depend on the country/region, socio-demographic factors, etc.

To properly annotate this nutritional data using a common language, the most relevant ontology in nutrition research is FoodOn (4). FoodOn is a comprehensive ontology composed of ”term hierarchy facets” that cover basic raw food source ingredients, packaging methods, cooking methods and preservation methods. It also includes an upper-level consisting of a variety of product type schemes under which food products can be categorized. On the other hand, the Metabolomics Standards Initiative has also highlighted the importance of ontologies in metabolomics (5). As Schlegel et al. reported, ”the application of ontologies to metabolomics can improve the consistency of study data and can help link data using relationships that extend the computational capacity of the study data and enrich that knowledge source with a myriad of nationally available data to help fuel hypothesis driven laboratory based research” (6).

In response to this, ChEBI (Chemical Entities of Biological Interest, https://www.ebi.ac.uk/chebi/) has developed a reference ontology for describing chemical compounds of biological interest in terms of their chemical structures, chemical categories and roles (7). The ChEBI ontology is manually maintained and annotated. More recently, an automatic method for describing and classifying chemicals, called ClassyFire (8), has been developed and widely adopted by databases such as ChEBI, PubChem (9) and the Human Metabolome Database (HMDB) (10). ClassyFire uses the ChemOnt ontology, consisting of more than 4800 different categories (with definitions) hierarchically structured into 11 different levels (Kingdom, SuperClass, Class, SubClass, etc.). Additionally, the HMDB has developed the ChemFOnt (chemical functional ontology) to describe the biological and industrial functions of all the compounds and metabolites found in this database. ChemFOnt consists of 4 major categories (physiological effect, disposition, process and role), 152 sub-categories and more than 4100 defined terms.

Although most existing ontologies have been specifically designed for a single theme, there are also some others composed of interconnected sub-ontologies, thus enabling users to establish relationships among different variables. For instance, ChEBI is organized in two sub-ontologies: 1) ”Molecular Structure”, in which molecular entities are classified according to structure and 2) ”Subatomic Particle”, which classifies particles smaller than atoms. On the other hand, the Gene Ontology, includes three independent sub-ontologies: 1) ”Biological process”, referred to a biological

objective to which the gene or gene product contributes; 2) ”Molecular function”, defined as the biochemical activity of a gene product; and 3) ”Cellular component”, which refers to the place in the cell where a gene product is active (11). In this regard, we would argue that nutritional research also generates large amounts of complex and inter-related data coming from self-reporting methods and metabolomics experiments. Therefore an interconnected set of sub-ontologies would be particularly useful for defining relationships between both metabolomics data and self-reported dietary questionnaires. To facilitate the construction of such an ontology that describes both foods and their associated metabolite biomarkers, we will draw from several open-access databases. These include Exposome-Explorer (12), Phenol-Explorer (13), PhytoHub (http://phytohub.eu/) and Food Database (FooDB) (http://foodb.ca/) - all of which contain rich information about food constituents and food metabolites.

However, relationships between foods and their metabolites are extremely complex and the way they are described varies tremendously across these databases. This lack of commonality and the lack of a common, hierarchical structure makes data comparison and data searching quite difficult. Therefore, the development of a comprehensive ontology to clearly define the relationships between nutritional (food composition) and metabolomics (food metabolite or biomarker) data is needed. This ontology could have multiple practical applications in nutrimetabolomics, being the annotation of terms using a consistent and standardized nomenclature the most basic one, but of great importance in this research field due to the inherent complexity and heterogeneity of the data managed (i.e. multiple names/synonyms to define the same food/metabolite). Additionally, other potential applications of the ontology could be the ability to perform different enrichment analysis (e.g. to investigate patterns of food consumption on the basis of metabolomics datasets) or to conduct semantic similarity analysis (e.g. to establish novel associations between foods and metabolites). In this work we describe FOBI (the Food-Biomarker Ontology), an ontology created with the aim of providing a common language to describe the many complex relationships in nutrimetabolomics research. This new ontology allow users (and online databases) to integrate dictionaries and analyze these two kinds of data independently or together in a consistent and homogeneous way.

FOBI is a freely available comprehensive ontology composed of two interconnected sub-ontologies including the ”Food Ontology” and the ”Biomarker Ontology”. This ontology is available in OWL (Web Ontology Language) and OBO (Open Biomedical Ontologies) formats at the project’s Github repository (https://github.com/pcastellanoescuder/FoodBiomarkerOntology). FOBI consists of 1184 terms, 4 different properties, 13 food top-level classes, 11 biomarker top-level classes and more than 4500 relationships. Furthermore, FOBI is part of OBOFoundry project and FOBI IDs have been indexed into the HMDB and FooDB databases to facilitate the interoperability and the exchange of data.

The FOBI’s “food sub-ontology” was created on the basis of dietary data obtained from self-reported surveys for dietary assessment, including food frequency questionnaires (FFQ) and dietary recalls (DRs). To expand this sub-ontology as much as possible, we adopted most of the entities from the FoodOn reference ontology.

Accordingly, the this sub-ontology is composed of more than 300 entities classified in different food classes. For this purpose, we considered both ”raw foods” and ”multi-component foods”, with a multi-component food defined as any food item composed by two or more raw foods. In turn, the food sub-ontology also describes the major ingredients forming part of each multi-component food according to the literature (14, 15). These entities were annotated using a common nomenclature to reduce the complexity and heterogeneity of dietary data collected from free text questionnaires.

Major food classes in the Food Ontology were created considering both the nature of the food and the availability of food intake biomarkers for each class. A total of 13 food top-level classes were generated: beverage food product, cacao food product, dairy food product, egg food product, flavouring additive, fruits and vegetables, grain plant, lipid food product, meat food product, multi-component food, nuts and legumes, spice or herb and sugar. In turn, each of these 13 top-level classes have different subclass structures depending on its nature.

On the other hand, food intake biomarkers (FIBs) are compounds derived directly from foods or the metabolism of food compounds that are characteristic or particular to a specific food item or food category. An important aspect to highlight on this regard is that, although the concentration of these metabolites in the food product may vary as a response to different factors (e.g. variety, agronomic practices, breeding, food processing), FIBs can always be associated with the consumption of the corresponding food (i.e. apple always contains phloretin, regardless the variety or cultivation conditions). FIBs potentially consist of a vast number of chemicals with very different physico-chemical properties, including polyphenols and carotenoids, coming from plant-derived foods; derivatives of amino acids and fatty acids (mainly found in animal products); methylxanthines from coffee, tea and cocoa; alkaloids, organic acids and many others. Food constituents can undergo multiple biotransformation steps after ingestion, thus significantly expanding their metabolic complexity. Typically xenobiotic food constituents are first subjected to phase I and phase II transformations, principally in the liver, kidneys and intestine, for detoxification purposes and to facilitate their excretion. Phase I metabolism normally involves cytochrome P450-mediated oxidation and hydrolysis transformations, while phase II reactions consist of chemical conjugations, such as methylation, acetylation, sulfation, glucuronidation and amino acid conjugation. The gut microbiota also plays a major role in the metabolism of poorly bioavailable food derived metabolites, usually involving ring cleavage reactions and a variety of fermentative pathways to produce smaller, more easily absorbed derivatives. Rather than trying to handle all possible compounds (possibly numbering in the tens of thousands) we chose to gather currently reported food derived metabolites and to define their relationships with foods and dietary patterns.

To create the "biomarker sub-ontology", we considered almost 600 known food metabolites, including dietary compounds and their host and microbiota-derived metabolites. These compounds were compiled from extensive literature reviews, and the information contained in open access databases such as Phenol-Explorer, PhytoHUB and the FooDB. Of particular help was the material produced by the EU-funded FoodBAll project (http://foodmetabolome.org/), which worked on discovering and validating FIBs for a range of foods. The FIBs in the Biomarker Ontology were classified according to their chemical classes using ClassyFire and ChemOnt (version 2.1). A key challenge in creating this sub-ontology was the complexity and diversity of the chemical nomenclature of food derived metabolites.

Besides the FOBI ID, this ontology also lists the code numbers for HMDB, KEGG, ChEBI, PubChem, InChIKey, InChI and ChemSpider for all these compounds, if available, which further facilitates the interoperability of FOBI and the exchange of data.

The architecture of FOBI is composed by classes corresponding to the items from the two sub-ontologies previously described (Food and Biomarker Sub-Ontologies), based on ChEBI (for metabolites) and FoodOn (for foods) respectively, and edges representing their relationships.

To conclude, FOBI is the first ontology that integrates nutritional and metabolomic data in a comprehensive common language. At the moment, FOBI has a total of 1184 terms (366 from Food Ontology and 818 from Biomarker Ontology), 11 chemical top-level classes, 13 food top-level classes and 4 different properties that are fully defined and which have clear relationship mappings. FOBI defines the relationships between foods and their metabolites (biomarkers) through a formal ontology.

FOBI allows experts to annotate and analyze nutritional and metabolomic data in a consistent way, making the results comparable between and across studies in the same field. The development of FOBI will lead to an improvement in the interoperability of nutritional and nutrimetabolomic data thereby making the data sets generated from these studies fully FAIR compliant.

References

1. Shim, J. S., Oh, K., & Kim, H. C. (2014) Dietary assessment methods in epidemiologic studies. Epidemiology and health, 36.
2. Scalbert, A., Brennan, L., Manach, C., et al. (2014) The food metabolome: a window over dietary exposure. The American Journal of Clinical Nutrition 99, 1286-1308.
3. Ulaszewska, M. M., Weinert, C. H., Trimigno, A., et al. (2019) Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies. Molecular nutrition & food research, 63, 1800384.
4. Dooley, D. M., Griffiths, E. J., Gosal, G. S., et al. (2018) FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. npj Science of Food, 2, 23.
5. Sansone, S. A., Schober, D., Atherton, H. J., et al. (2007) Metabolomics standards initiative: ontology working group work in progress. Metabolomics, 3, 249-256.
6. Schlegel, D. R., Ruttenberg, A., & Elkin, P. L. (2015) Ontologies in Metabolomics. Metabolomics, 5.
7. Degtyarenko, K., De Matos, P., Ennis, M., et al. (2007) ChEBI: a database and ontology for chemical entities of biological interest. Nucleic acids research, 36, D344-D350.
8. Feunang, Y. D., Eisner, R., Knox, C., et al. (2016) ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. Journal of cheminformatics, 8, 61.
9. Kim, S., Chen, J., Cheng, T., et al. (2019) PubChem 2019 update: improved access to chemical data. Nucleic acids research, 47, D1102-D1109.
10. Wishart, D. S., Feunang, Y. D., Marcu, A., et al. (2017) HMDB 4.0: the human metabolome database for 2018. Nucleic acids research, 46, D608-D617.
11. Ashburner, M., Ball, C. A., Blake, J. A., et al. (2000) Gene ontology: tool for the unification of biology. Nature genetics, 25, 25.
12. Neveu, V., Moussy, A., Rouaix, H., et al. (2016) Exposome-Explorer: a manually-curated database on biomarkers of exposure to dietary and environmental factors. Nucleic acids research, gkw980.
13. Rothwell, J. A., Perez-Jimenez, J., Neveu, V., et al. (2013) Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. Database, 2013.
14. McCance, R. A., Widdowson, E. M. (2014) The Composition of Foods. Royal Society of Chemistry, Cambridge (UK).
15. Reinivuo, H., Bell, S., & Ovaskainen, M. L. (2009) Harmonisation of recipe calculation procedures in European food composition databases. Journal of Food Composition and Analysis, 22, 410-413.