Proteomics and Metabolomics: The Final Frontier of Nutrition Research

Richard D Semba
Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Introduction

Revolutionary new technologies allow us to penetrate scientific frontiers and open vast new territories for discovery. In astronomy, the Hubble Space Telescope has facilitated an unprecedented view outwards, beyond our galaxy. Wherever the telescope is directed, scientists are making exciting new observations of the deep universe. Another revolution is taking place in two fields of "omics" research: proteomics and metabolomics. In contrast, this view is directed inwards, towards the complexity of biological processes in living organisms. Proteomics is the study of the structure and function of proteins expressed by an organism. Metabolomics is the study of small, low molecular weight metabolites and their cellular processes. The study of individual proteins and metabolites has a long tradition, but a collective approach to their study developed only recently. The terms "proteome" and "metabolome" were first mentioned in the published scientific literature in 1996[1] and 1998,[2] respectively. The German botanist Hans Winkler (1877–1945) coined the related term "genome" more than seven decades earlier.[3]

Proteomics and metabolomics in the post-genomic era

In 2003, the Human Genome Project – which had the goal of mapping all the genes of the human genome – was declared complete.[4] In the post-genomic era, two major challenges in the life sciences include the elucidation of all the proteins and metabolites in the human body. The proteome and metabolome have a level of complexity that far exceeds the genome. In humans, ~20,000 protein-coding genes give rise to ~100,000 proteins and an estimated one million different protein modified forms.[5,6] The many forms of proteins arise from mutations, RNA editing, RNA splicing, post-translational modifications, and protein degradation; the proteome does not strictly reflect the genome. Proteins function as enzymes, hormones, receptors, immune mediators, structure, transporters, and modulators of cell communication and signaling. The metabolome consists of amino acids, amines, peptides, sugars, oligonucleotides, ketones, aldehydes, lipids, steroids, vitamins, and other molecules. These metabolites reflect intrinsic chemical processes in cells as well as environmental exposures such as diet and gut microbial flora. The current Human Metabolome Database contains more than 40,000 entries[7] – a number that is expected to grow quickly in the future.

The goals of proteomics include the detection of the diversity of proteins, their quantity, their isoforms, and the localization and interactions of proteins. The goals of metabolomics include mapping the function of metabolic pathways, many of which remain partially or completely uncharacterized,[8] as well as detecting and measuring the diversity and dynamic changes of metabolites. This fundamental work should help lead to the discovery of new biological mechanisms, biomarkers, drug targets, and pathways of disease. Proteomics and metabolomics are vital steps in the progress of science towards translational research, clinical trials, and personalized medicine. The research fields of cancer, neurology, endocrinology, and cardiovascular disease have been in the vanguard in using proteomic and metabolomic approaches in scientific investigation. In contrast, the field of nutrition has been slow in applying these powerful techniques.

> "Proteomics and metabolomics are vital steps in the progress of science"

The technology to investigate the immense complexity of the proteome and metabolome has advanced rapidly within the last several years. Newer mass spectrometers have greater sensitivity, higher reproducibility, better comprehensiveness, and more rapid throughput, allowing the identification and quantification of thousands of proteins and metabolites in tissues and samples. Mass spectrometers are at the heart of the laboratory (**Textbox 1**). The Orbitrap mass analyzer was commercially available in 2005[9] and gave rise to subsequent generations of Orbitrap mass spectrometers.[10–12] The comprehensive analysis of proteins and lipids has been increased dramatically by sequential windowed data-independent acquisition of the total high resolution mass spectra (SWATH-MS) on triple time-of-flight mass spectrometers.[13,14] Multiple reaction monitoring (MRM; also known as selected reaction monitoring), a targeted mass spectrometry technique, can use triple quadrupole, or QTRAP™, mass spectrometers to measure sets of proteins.[15] MRM allows precise, antibody-free quantitation of proteins in a multiplexed fashion. Targeted metabolomic approaches can be conducted using QTRAP™.[16,17] Nuclear magnetic resonance spectroscopy can also be used to measure metabolites, with the advantage of minimal sample preparation but the major disadvantage of low sensitivity.[18]

Technological breakthroughs are accelerating the research

Currently it is possible to measure thousands of proteins in biological samples – something that was not feasible several years ago.[19] These advances are due not only to the new innovations in mass spectrometry, but also to improvements in sample preparation such as depletion of highly abundant proteins, thus allowing detection of low abundance proteins (the "deep proteome"),[20] more effective electrophoresis and chromatography protocols, and better tools for quantification (**Textbox 2**).[21–23] Protocols are also improving for detection of post-translation modifications (PTMs), such as phosphorylation, glycosylation, acetylation, ubiquitination and sumoylation. PTMs are important to study since they reflect the diversity of protein function. Many of the PTMs are difficult to study because they are labile to sample processing and mass spectrometry. For example, O-GlcNAcylation, an important PTM that rivals phosphorylation in abundance and distribution, has been especially challenging to detect and measure.[24] Many proteins have functions that are unknown or not well understood. By studying the proteins with which a particular protein interacts, it is possible to deduce biological functions and pathways.[25] Protocols have recently been developed for proteomic analysis of dried blood spots[26] and formalin-fixed, paraffin-embedded tissues.[27]

Textbox 1

What is a mass spectrometer?

A mass spectrometer is simply an instrument that weighs molecules. Mass spectrometry is a technique that separates charged molecular species based upon their mass-to-charge ratio (m/z). The instrument provides a mass spectrum, which gives:

(1) The number of components;
(2) m/z for each component;
(3) sequence information (tandem mass spectrometry); and
(4) abundance information.

Spectra are used to identify specific molecules and their relative amounts. There are many different kinds of mass spectrometers. Most laboratories involved in protein biomarker discovery studies or investigations of protein interactions have at least two mass spectrometers for liquid chromatography-tandem mass spectrometry (LC-MS/MS), an Orbitrap or TripleTOF for discovery work and a QTRAP™ for targeted quantitation of proteins.

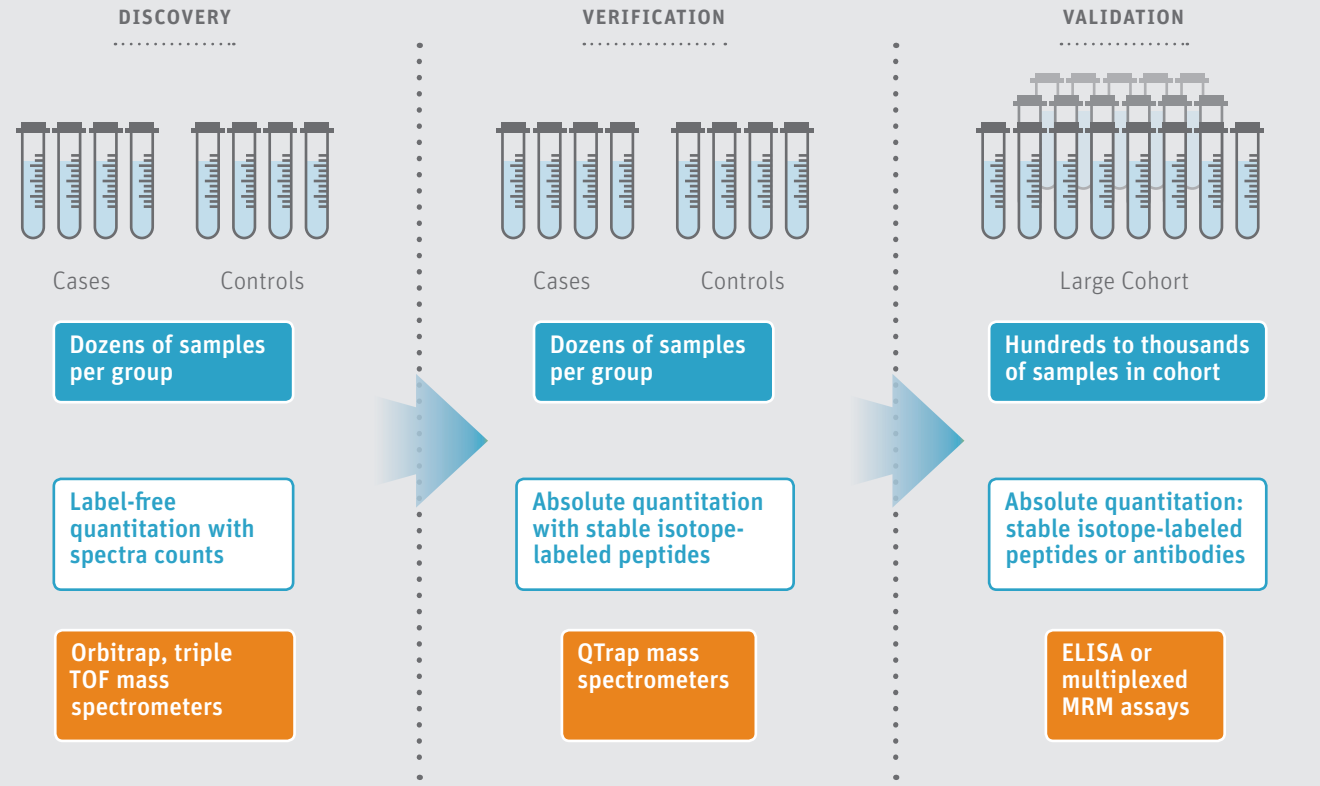
Textbox 2

Key messages

- The diversity of the proteome and metabolome are far greater than the genome, and probably hold the keys to understanding the biology of health and disease.
- Recent advances in sample preparation, instrumentation, and bioinformatics have revolutionized proteomics and metabolomics.
- It is now possible to measure thousands of proteins and hundreds of metabolites in biological samples – something that was not possible even several years ago.
- Proteomics and metabolomics are entering a large growth phase and offer great opportunities for young investigators.

Many recent metabolomic studies have used so-called "targeted" approaches in which a panel of well-characterized and validated metabolites is measured using liquid chromatography-mass spectrometry (LC-MS).[28–30] Given the complexity of various metabolites, there is no single analytical platform to

FIGURE 1: An example of a workflow using a proteomic approach to identify specific biomarkers and pathways in plasma samples



In the hypothesis-free discovery phase, a case-control design is used to identify dozens to hundreds of proteins that are higher or lower in concentrations between the two groups. The verification phase involves absolute quantitation of the candidate proteins between the two groups using stable isotope labeled standard peptides. The proteins that are verified require further validation in large cohort studies. ELISAs or multiple reaction monitoring assays (MRM) can be devised for measuring these specific proteins. Workflows may vary depending upon the clinical phenotype under investigation.

measure the complete metabolome. At present, about 200–1,000 metabolites can be measured using LC-MS in commercial labs or some academic labs, depending upon the type of sample.

Bioinformatics has played a vital role in the acceleration of proteomics and metabolomics. Raw MS data from proteomic analyses can be analyzed using open source search engines such as X!Tandem, OMSSA, or proprietary databases such as Mascot and Sequest. The software assigns sequence information for peptides based upon the spectra, and then protein identifications based upon the specific peptides. Authoritative and comprehensive protein databases include neXtProt[5] for human proteins and UniProt.[31] Annotated databases such as Gene Ontology (GO)[32] and pathway databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG)[33] and Database for Annotation, Visualization and Integrated Discovery (DAVID)[34] are particularly useful. Online resources and databases of metabolites include Metabolomics Workbench, METLIN, and BiGG.[35]

> "Bioinformatics has played a vital role in the acceleration of proteomics and metabolomics"

Phases of investigation in biomarker discovery

Proteomics and metabolomics are used to discover new biological pathways and biomarkers that are associated with specific diseases or conditions. A typical workflow for the identification of novel protein biomarkers proceeds in two phases of investigation: a hypothesis-free discovery phase followed by verification and validation phases (**Figure 1**). The discovery phase may typically utilize a case-control study design with at least thirty subjects per group. The two groups are compared in order to determine which proteins or metabolites are differentially expressed between the two groups so as to identify potential candidate biomarkers. In the case of proteomics, the comparison of

FIGURE 2: Examples of mass spectrometers and instrumentation for proteomic and metabolomic studies, author's laboratory team

a. We use a 5600+ TripleTOF (AB Sciex) mass spectrometer for discovery phase proteomic and lipidomic studies. Dr Alexey Lyashkov operating the instrument.

b. For targeted proteomic and metabolic studies, we use a 5500 QTrap (AB Sciex) mass spectrometer. Dr Pingbo Zhang checking the instrument.

c. A laboratory automated robotic workstation allows efficient, accurate, and high throughput sample handling. Beckman Coulter Biomek NX^{P} robot.

protein concentrations is usually based upon label-free methods, such as spectral counts. Further verification is needed using precise gold standard quantitation measurements such as MRM using stable isotope-labeled peptide standards and LC-MS/MS on a QTRAP™. It is possible to measure several dozen different plasma proteins in a single multiplexed assay using MRM.[36,37]

For proteins, the validation phase involves applying MRM or devising more conventional assays, such as enzyme-linked immunosorbent assays (ELISA) for measurement of specific proteins, and then measuring the biomarkers across one or two large cohort studies. The case-control study design, with careful attention to selection of cases and controls[35,38] and sample size and power,[39] provides a rigorous strategy for discovering new biological pathways, biomarkers, and therapeutic targets. The choice of a well-defined clinical phenotype is critical in the study design of proteomic or metabolomic biomarker studies. For example, bone density assessed by dual-energy X-ray absorptiometry would be preferable to plasma 25-hydroxyvitamin D concentrations as a clinical outcome measure in older adults. Quality standards have been developed for reporting information from proteomic studies.[40,41] Since LC-MS assays for metabolites provide absolute quantitation, discovery/verification can be followed by validation in large cohort studies.

There was great enthusiasm and expectations a decade ago about the potential of proteomics and metabolomics to lead to the discovery of new circulating biomarkers. However, these fields did not deliver immediate returns in this earlier period because of existing limitations in study design, sample preparation, mass spectrometry instrumentation, and bioinformatics. Even the term "biomarker" became viewed in a slightly jaded light. Some notable successes are beginning to emerge, such as validated plasma protein biomarkers for hematopoietic stem cell transplantation,[42] and 2-aminoadipic acid as a predictor of type 2 diabetes.[43] It should be emphasized that proteomics and metabolomics are not just about biomarkers. These fields are providing fundamental knowledge of cellular processes, molecular interactions, and localization of biological pathways.

> "Proteomics and metabolomics are providing fundamental knowledge of cellular processes, molecular interactions, and localization of biological pathways"

Current challenges in proteomics and metabolomics

There are three practical challenges to using proteomic and metabolomic approaches in research.

First, there are a limited number of laboratories with the scientific expertise and instrumentation to conduct these studies. Mass spectrometry core facilities in some universities may be able to meet the needs of nutrition investigators. Some investi-

gators have built up their own labs for proteomic and metabolomic investigations (**Figure 2**).

Second, it is very expensive and time-consuming to prepare samples and conduct sample analyses using LC-MS/MS on mass spectrometers. In most laboratories, mass spectrometers are running 24 hours a day, operated by highly specialized experts who have many years of experience in dealing with these complicated instruments. Core labs generally charge high rates for instrument time, given the great cost for the acquisition and maintenance of mass spectrometers.

Third, proteomic and metabolomic investigations often yield a bewildering wealth of data that require expertise in bioinformatics and biostatistics. The immense complexity of data is often perceived as a "bottleneck" in proteomics and metabolomics research. There are a growing number of specialists with training in bioinformatics. Bioinformatic software, search engines, and methodologies are evolving rapidly. When the final results are placed in the hands of the investigator, the ultimate and most important challenge is in the interpretation of the proteins and metabolites that have been identified in the study. The investigator should learn to "expect the unexpected." The application of such powerful tools may uncover novel biological pathways and relationships – exciting discoveries that can move the field forward.

How does one get started in proteomics or metabolomics? Institutions such as the Cold Spring Harbor Laboratories, the Wellcome Trust, and the Seattle Proteome Center offer specialized courses in proteomics. The European Bioinformatics Institute and West Coast Metabolomics Center give courses in metabolomics. Many universities now have regular courses in proteomics and metabolomics. The main international forum for proteomics research is the Human Proteome Organization (HUPO), which was founded in 2001. The main goal of HUPO is to identify and characterize the diversity of human proteins.[44] HUPO has specific disease and organ-specific proteomic initiatives, such as brain,[45] liver,[46] and eye.[47] The Metabolomics Society, incorporated in 2004, promotes the application of metabolomics in the life sciences. Although proteomics and metabolomics are at currently at the frontier of nutrition research, the fields are on the ascendancy and should provide unprecedented opportunities for young investigators. Some nutrition research groups have already taken up the challenge.[48-50] As noted by the German scientist Georg Christoph Lichtenberg (1742–1799): "Where the frontier of science once was is now the center."

Correspondence: ***Richard D Semba,***
Wilmer Eye Institute, Johns Hopkins School of Medicine, Smith Building, M015, 400 N. Broadway, Baltimore, MD 21087, USA.
Email: *rdsemba@jhmi.edu*

References

01. Wilkins MR, Pasquali C, Appel RD et al. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Biotechnology (NY) 1996;14:61–5.
02. Tweeddale H, Notley-McRobb L, Ferenci T. Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ("metabolome") analysis. J Bacteriol 1998;180:5109–116.
03. Winkler H. Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche. Jena: Verlag Fischer, 1920.
04. Collins FS, Green ED, Guttmacher AE et al. A vision for the future of genomics research: a blueprint for the genomic era. Nature 2003;422:835–47.
05. Gaudet P, Michel PA, Zahn-Zabal M et al. The neXtProt knowledgebase on human proteins: current status. Nucleic Acids Res 2015;43:D764–70.
06. Fu W, O'Connor TD, Jun G et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 2013;493:216–20.
07. Wishart DS, Jewison T, Guo AC et al. HMDB 3.0 – The Human Metabolome Database in 2013. Nucleic Acids Res 2013;41:D801–7.
08. Mulvihill MM, Nomura DK. Metabolomic strategies to map functions of metabolic pathways. Am J Physiol Endocrinol Metab 2014;307:E237–44.
09. Hu Q, Noll RJ, Makarov A et al. The Orbitrap: a new mass spectrometer. J Mass Spectrom 2005;40:430–43.
10. Michalski A, Damoc E, Lange O et al. Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. Mol Cell Proteomics 2012;11:O111.013698.
11. Michalski A, Damoc E, Hauschild JP et al. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol Cell Proteomics 2011;10:M111.011015.
12. Almeida R, Pauling JK, Sokol E et al. Comprehensive lipidome analysis by shotgun lipidomics on a hybrid quadrupole-orbitrap-linear ion trap mass spectrometer. J Am Soc Mass Spectrom 2015;26:133–48.
13. Gillet LC, Navarro P, Tate S et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics 2012;11:10.1074/mcp.O111.016717.
14. Simons B, Kauhanen D, Sylvänne T et al. Shotgun lipidomics by sequential precursor ion fragmentation on a hybrid quadrupole time-of-flight mass spectrometer. Metabolites 2012:20:195–213.
15. Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls, and future directions. Nat Methods 2012;9:555–66.
16. Illig T, Gieger C, Zhai G et al. A genome-wide perspective of genetic variation in human metabolism. Nat Genet 2010;52:137–41.
17. Menni C, Zhai G, Macgregor A et al. Targeted metabolomics profiles

are strongly correlated with nutritional patterns in women. Metabolomics 2013;9:506–14.

18. Emwas AH. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. Methods Mol Biol 2015;1277:161–93.
19. Van Eyk JE. The maturing of proteomics in cardiovascular research. Circ Res 2011;108:490–8.
20. Shi T, Zhou JY, Gritsenko MA et al. IgY14 and SuperMix immunoaffinity separations coupled with liquid chromatography-mass spectrometry for human plasma proteomics biomarker discovery. Methods 2012;56:246–53.
21. Megger DA, Bracht T, Meyer HE et al. Label-free quantification in clinical proteomics. Biochim Biophys Acta 2013;1834:1581–90.
22. Evans C, Noirel J, Ow SY et al. An insight into iTRAQ: where do we stand now? Anal Bioanal Chem 2012;404:1011–27.
23. Hoedt E, Zhang G, Neubert TA. Stable isotope labeling by amino acids in cell culture (SILAC) for quantitative proteomics. Adv Exp Med Biol 2014;806:93–106.
24. Hart GW. Three decades of research on O-GlcNAcylation – a major nutrient sensor that regulates signaling, transcription and cellular metabolism. Front Endocrinol (Lausanne) 2014;5:183.
25. Miteva YV, Budayeva HG, Cristea IM. Proteomics-based methods for discovery, quantification, and validation of protein-protein interactions. Anal Chem 2013;85:749–68.
26. Martin NJ, Bunch J, Cooper JH. Dried blood spot proteomics: surface extraction of endogenous proteins coupled with automated sample preparation and mass spectrometry analysis. J Am Soc Mass Spectrom 2013;24:1242–9.
27. Sprung RW Jr, Brock JWC, Tanksley JP et al. Equivalence of protein inventories obtained from formalin-fixed paraffin-embedded and frozen tissue in multidimensional liquid chromatography-tandem mass spectrometry shotgun proteomic analysis. Mol Cell Proteomics 2009;8:1988–98.
28. Illig T, Gieger C, Zhai G et al. A genome-wide perspective of genetic variation in human metabolism. Nat Genet 2010;52:137–41.
29. Mapstone M, Cheema AK, Fiandaca MS et al. Plasma phospholipids identify antecedent memory impairment in older adults. Nat Med 2014;20:415–8.
30. Barnes VM, Kennedy AD, Panagakos F et al. Global metabolomic analysis of human saliva and plasma from healthy and diabetic subjects, with and without periodontal disease. PLoS One. 2014;9:e105181.
31. UniProt Consortium. Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res 2013;41:D43–7.
32. Ashburner M, Ball CA, Blake JA et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.
33. Kanehisa M, Goto S, Kawashima S et al. The KEGG resource for deciphering the genome. Nucleic Acids Res 2004;32:D277–80.
34. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57.
35. Tzoulaki I, Ebbels TMD, Valdes A et al. Design and analysis of metabolomics studies in epidemiologic research: a primer on –omic technologies. Am J Epidemiol 2014;180:129–39.
36. Kuzyk MA, Smith D, Yang J et al. Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. Mol Cell Proteomics 2009;8:1860–77.
37. Domanski D, Percy AJ, Yang J et al. MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. Proteomics 2012;12:1222–43.
38. Rundle A, Ahsan H, Vineis P. Better cancer biomarker discovery through better study design. Eur J Clin Investig 2012;42:1350–9.
39. Skates SJ, Gillette MA, LaBaer J et al. Statistical design for biospecimen cohort size in proteomics-based biomarker discovery and verification studies. J Proteome Res 2013;12:5383–94.
40. Taylor CF, Paton NW, Lilley KS et al. The minimum information about a proteomics experiment (MIAPE). Nat Biotech 2007;25:887–93.
41. Deutsch EW, Albar JP, Binz PA et al. Development of data representation standards by the Human Proteome Organization Proteomics Standards Initiative. J Am Med Inform Assoc 2015; Feb 28 [Epub ahead of print].
42. Paczesny S, Duncan C, Jacobsohn D et al. Opportunities and challenges of proteomics in pediatric patients: circulating biomarkers after hematopoietic stem cell transplantation as a successful example. Proteomics Clin Appl 2014;8:837–50.
43. Wang TJ, Ngo D, Psychogios N et al. 2-Aminoadipic acid is a biomarker for diabetes risk. J Clin Invest 2013;123:4309–17.
44. Omenn GS. The strategy, organization, and progress of the HUPO Human Proteome Project. J Proteomics 2014;100:3–7.
45. Hamacher M, Stephan C, Eisenacher M et al. Maintaining standardization: an update of the HUPO Brain Proteome Project. Expert Rev Proteomics 2008;5:165–73.
46. Gao X, Zheng J, Beretta L et al. The 2009 Human Liver Proteome Project (HLPP) Workshop 26 September 2009, Toronto, Canada. Proteomics 2010;10:3058–61.
47. Semba RD, Enghild JJ, Venkatraman V et al. The Human Eye Proteome Project: perspectives on an emerging proteome. Proteomics 2013;13:2500–11.
48. Cole RN, Ruczinski I, Schulze K et al. The plasma proteome identifies expected and novel proteins correlated with micronutrient status in undernourished Nepalese children. J Nutr 2013;143:1540–8.
49. Walker GE, Ricotti R, Roccio M et al. Pediatric obesity and vitamin D deficiency: a proteomic approach identifies multimeric adiponectin as a key link between these conditions. PLoS One 2014;9:e83685.
50. Andersen MB, Rinnan Å, Manach C et al. Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. J Proteome Res 2014;13:1405–18.