

## LIPIDOMICS AND LIPID BIOMARKER DISCOVERY

Peter Meikle\*, Christopher Barlow and Jacqui Weir

Metabolomics Laboratory, Baker IDI Heart and Diabetes Institute, VIC 3004

\*Corresponding author: [peter.meikle@bakeridi.edu.au](mailto:peter.meikle@bakeridi.edu.au)

### Introduction

As the metabolome can be defined as the small molecule (metabolite) complement of a biological system, so the lipidome can be defined as the lipid complement of the same biological system. Lipidomics then is the systematic study of the lipidome and as with metabolomics, lipidomics has its own unique advantages and challenges.

There are many thousands of different lipid structures within biological systems. While the theoretical goal of lipidomics may be the analysis of the complete lipid complement of the system under investigation, there are a number of technical and practical issues that limit this approach. As a result, lipidomic studies (outlined in Fig. 1) are restricted to a subset of lipids that may reflect abundance, specific lipid classes of interest or structural characteristics.

Lipidomics studies are often focussed on either the identification and validation of lipid biomarkers or the characterisation of lipid metabolism with a view to understanding the system/disease under investigation. This review will discuss the technical issues related to lipidomic studies as well as the different approaches to mapping the lipidome and identifying lipid biomarkers.

### Lipids: Classification, Structure and Distribution

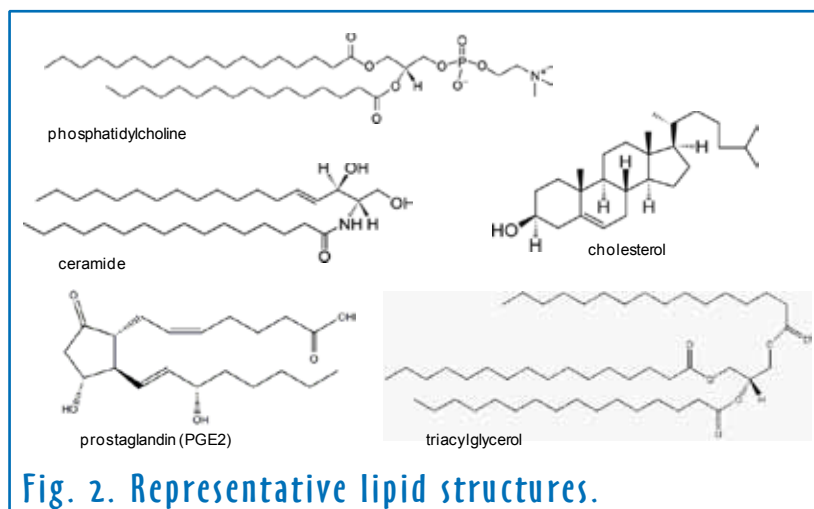
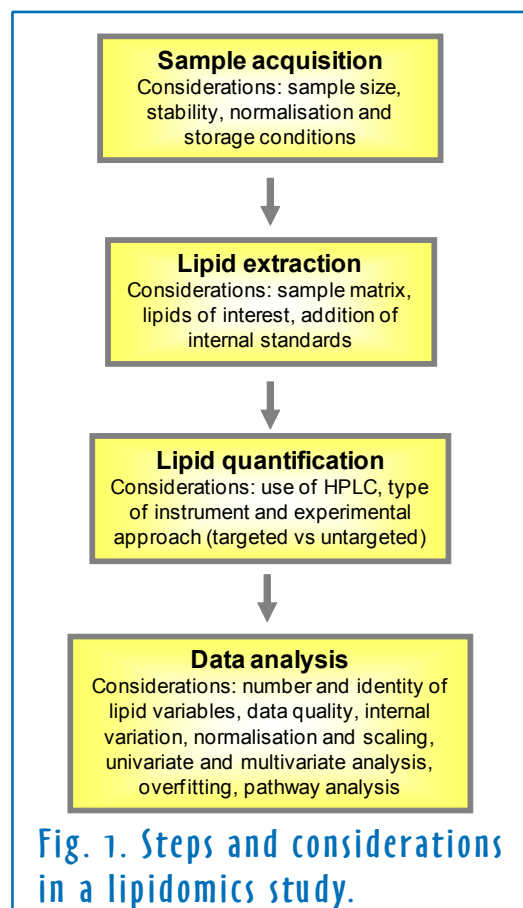
Lipids are a diverse group of small molecules which, broadly speaking, are hydrophobic or amphiphilic in nature. There have been a number of classification schemes proposed; however, most recently, members of the LIPID MAPS (LIPID Metabolites And Pathways Strategy) consortium have proposed a scheme that deals with the enormous diversity of lipid structures and is compatible with the informatics requirements of the lipidomics era (1). In this system, lipids are classified into eight groups (Table 1) with subsequent sub-classification groups in each of these.

The diverse nature of lipid structures (Fig. 2) and their properties enable these compounds to fulfill many biological functions from energy storage through membrane structure to signaling intermediates. However, this same complexity poses many challenges to the characterisation, identification and quantification of this diverse class of molecules. A number of websites provide useful overviews of lipid structure and function as well as detailed protocols for lipid extraction and separation by HPLC. See:

LIPID MAPS <http://www.lipidmaps.org>  
 Lipid Library <http://lipidlibrary.co.uk>  
 Lipid Bank <http://lipidbank.jp>  
 Cyberlipids <http://www.cyberlipid.org>

### Lipid Analysis

**Sampling and lipid extraction:** Before sampling, consideration of an accurate method of sample normalisation is required; this could include volume, protein content, cell number or weight of tissue. The method of extraction should also be compatible with the sample matrix, the lipids under investigation and the subsequent analytical method to be used. The most widely accepted method for extraction of lipids is the method of



**Table 1. Lipid categories.**

Category	Abbreviation	Example
Fatty acyls	FA	Fatty acids and eicosanoids
Glycerolipids	GL	Diacyl- and triacylglycerols
Glycerophospholipids	GP	Phosphatidylcholines and phosphatidylethanolamines
Sphingolipids	SP	Ceramides, glycolipids and sphingomyelins
Sterol lipids	ST	Cholesterol, bile acids and sterol hormones
Prenol lipids	PR	Isoprenoids
Saccharolipids	SL	Acylaminosugars
Polyketides	PK	Flavanoids

Folch (2). The original method described the extraction of hydrophobic and hydrophilic lipids into a two-phase system consisting of chloroform, methanol and water. Subsequent modifications were made to accommodate larger volumes or to select for specific lipid groups by altering pH. While these methods provide good recoveries of most lipid species in the organic phase and remove the majority of salts, some lipids, such as gangliosides, will partition primarily into the aqueous phase and will require additional chromatography for full recovery.

As a result of the high sensitivity displayed by modern instruments, the sample size required for lipidomic studies has decreased considerably and this has enabled the use of alternate extraction procedures. For example, our laboratory uses a modification of the Folch method whereby there is no partitioning between water and organic phases. This allows the rapid extraction of the full range of lipids (typically we use 5  $\mu$ L plasma or 100  $\mu$ g wet weight cells/tissue). Further information on tailoring extraction methods can be found at the websites listed above.

**Internal standards:** Quantification in lipidomics is usually achieved by comparison with internal standards added to the biological matrix prior to lipid extraction. The choice of suitable internal standards is therefore critical if accurate quantification is to be achieved. Ideally, internal standards should have similar chemical and physical properties to the lipids to which they are being compared in order to minimise any difference in extraction efficiency or detection between the two. Stable-isotope labelled lipids represent an excellent choice for internal standards as they are chemically indistinguishable from their non-labelled isotopologues. However, the large number of lipids examined in many experiments makes it infeasible to match every lipid with a single internal standard, and so in practice, one or two internal standards are used for a particular class of lipids. An alternative to stable-isotope labelled standards is to use non-physiological lipids belonging to the same class; most commonly, this involves the incorporation of an odd-chain fatty acid.

**High pressure liquid chromatography (HPLC):** HPLC coupled to mass spectrometry (LC-MS) allows for the partial separation of lipids prior to MS analysis; this is particularly useful when the goal is a full analysis of complex lipid mixtures. When selecting HPLC parameters such as solvent composition and type of column, the nature

of the lipids of interest should be considered carefully. Recent advances in column and instrument technology with respect to particle size, bonded phases and pressure limits have afforded a wide range of columns available for specific uses. Lipidomics studies usually use reverse phase columns with solvent gradients of increasing organic composition. Suitable chromatographic conditions have been reported for most lipid classes; however, conditions that are appropriate for a broad range of lipids are less common. In our hands, the tetrahydrofuran:methanol:water system reported by Yoo *et al.* (3) works well for most lipid classes. For analysis of individual lipid classes, a normal phase column or an alternative solvent system may be more appropriate (4,5; see also websites listed above).

**Mass spectrometry:** Quantification of lipids is typically performed using mass spectrometry. Mass spectrometry is a technique that measures the mass to charge ratio ( $m/z$ ) of gas phase ions. Ionisation of most lipid species results in singly charged ions and, consequently, mass spectrometry provides a convenient measure of mass. Prior to the 1980s, mass spectrometry was limited to the analysis of volatile analytes by electron impact or chemical ionisation, and consequently, its use in lipidomics was relatively restricted or required extensive sample derivatisation to increase volatility. Gas chromatography-mass spectrometry (GC-MS) is still a powerful method for fatty acid analysis (6). The development of new soft ionisation techniques such as electrospray ionisation (ESI) and matrix-assisted laser desorption/ionisation (MALDI), for which Fenn and Tanaka shared in the 2002 Nobel Prize for Chemistry, enabled the introduction of non-volatile species into the gas phase and thereby made mass spectrometry applicable to a wide array of biomolecules including lipids. ESI-mass spectrometry (ESI-MS) is particularly important in lipidomics due to its high sensitivity, small sample size requirement and ability to be directly coupled to liquid chromatography. Furthermore, with suitable internal standards, ESI-MS allows quantitative analysis over several orders of magnitude and can routinely achieve limits of detection in the femtomolar range. Although ESI can generate gas-phase ions of most lipids, it should be emphasised that ionisation efficiency (and therefore detection limits) varies considerably depending on their physicochemical properties.

Given the large number of lipid structures possible, many of which are isomeric, mass alone is insufficient to uniquely

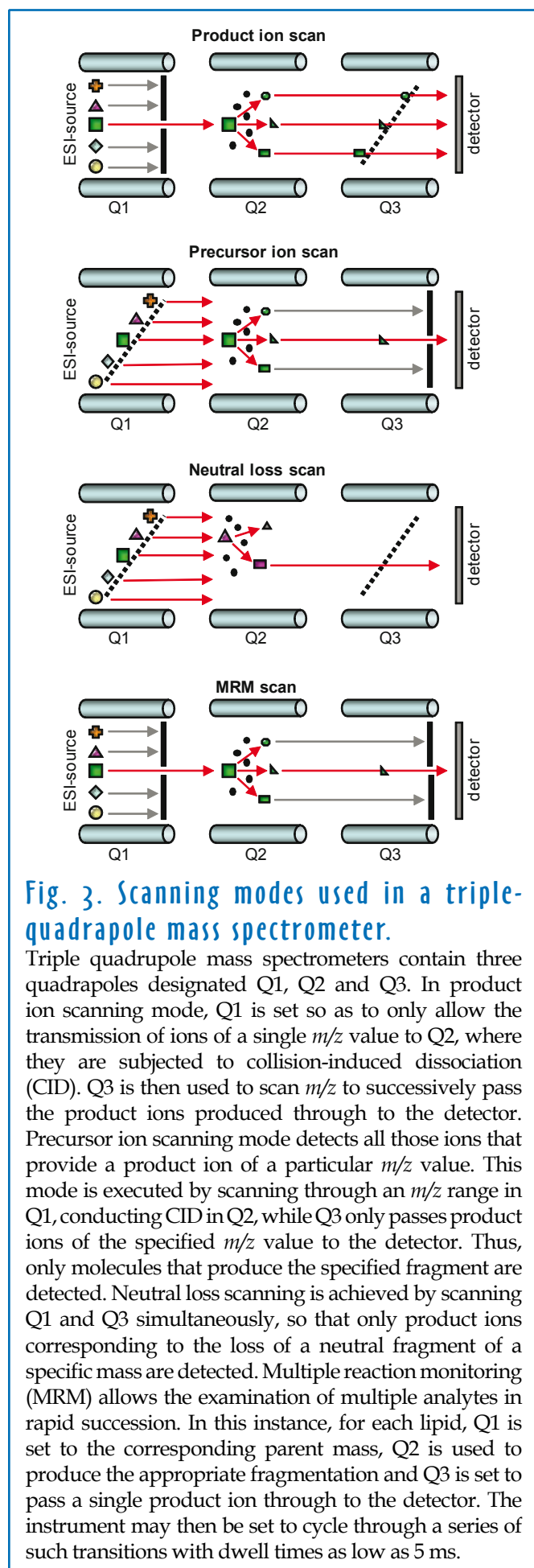
identify most lipids. Fortunately, mass spectrometry provides the opportunity to manipulate ions in the gas phase. In particular, fragmentation of selected ions, typically by collision-induced dissociation (CID) with an inert bath gas, can provide additional structural information about the selected ion, based either on the product ion or neutral fragment lost (7). CID is typically conducted using triple quadrupole instruments where ions pass through three successive quadrupoles (Q1, Q2 and Q3). The instrument may be operated in product ion, precursor ion or neutral loss scanning modes (Fig. 3). Most major lipid sub-groups display characteristic fragmentation patterns (Table 2), which facilitates their rapid analysis using precursor or neutral loss scanning experiments. Several precursor ion and neutral loss scans may be combined in a multi-dimensional fingerprinting approach to lipidomics (8). A more targeted approach detects specific lipids by using multiple reaction monitoring (MRM) mode (Fig. 3). MRM mode is of particular importance when on-line liquid chromatography is being utilised, as the chromatographic timescale is often too fast to allow multiple neutral loss or precursor ion scans. The use of MRM in conjunction with liquid chromatography may be further enhanced by only monitoring each MRM transition at its corresponding retention time (scheduled MRM). Under this scenario, it is possible, with suitable chromatographic conditions, to monitor several hundred lipid species in a single LC-MS analysis.

CID combined with the judicious choice of ESI polarity and choice of counter-ions can provide significant structural information (7,9,10), and in many cases, absolute identification of the lipid of interest is possible. These techniques have been used by Han and Gross to develop 'shotgun lipidomics', which combines CID and intra-source separation to facilitate the quantitative analysis of hundreds of lipid species (8,11,12). Numerous other approaches using ESI-quadrupole-time of flight (ESI-Q/TOF) (13) or MALDI-TOF (14) have also been developed. Each method is a compromise of sensitivity, accuracy, the number and type of lipids analysed, and the analysis time (sample throughput). The selection of the appropriate methodology for a given lipidomics study will require consideration of all these factors.

## Data Analysis

In many lipidomic studies, the primary objective is to identify and quantify phenotypic variation between study groups and to relate those differences back to the source of variation. Study groups may include cell culture models or biological samples from animal models and human cohorts. The source of variation may relate to differences in genetic background, environment (diet, exercise), disease (obesity, diabetes, heart disease) or treatment (therapeutics).

Lipidomics data can be produced as absolute quantification, semi-quantification or relative quantification of known lipid species, or as relative signal intensity of unknown lipid species represented by retention times and  $m/z$ . The analysis of lipidomic data is dependent on a number of factors, including the number and identity of lipid variables, the data quality, the sample size and the internal variation within sample groups. Preprocessing of data may require normalisation and scaling procedures (15).



**Fig. 3. Scanning modes used in a triple quadrupole mass spectrometer.**

Triple quadrupole mass spectrometers contain three quadrupoles designated Q1, Q2 and Q3. In product ion scanning mode, Q1 is set so as to only allow the transmission of ions of a single  $m/z$  value to Q2, where they are subjected to collision-induced dissociation (CID). Q3 is then used to scan  $m/z$  to successively pass the product ions produced through to the detector. Precursor ion scanning mode detects all those ions that provide a product ion of a particular  $m/z$  value. This mode is executed by scanning through an  $m/z$  range in Q1, conducting CID in Q2, while Q3 only passes product ions of the specified  $m/z$  value to the detector. Thus, only molecules that produce the specified fragment are detected. Neutral loss scanning is achieved by scanning Q1 and Q3 simultaneously, so that only product ions corresponding to the loss of a neutral fragment of a specific mass are detected. Multiple reaction monitoring (MRM) allows the examination of multiple analytes in rapid succession. In this instance, for each lipid, Q1 is set to the corresponding parent mass, Q2 is used to produce the appropriate fragmentation and Q3 is set to pass a single product ion through to the detector. The instrument may then be set to cycle through a series of such transitions with dwell times as low as 5 ms.

Univariate analysis can provide an initial measure of variation between groups and identify those individual lipid species contributing to that variation. Depending on the study groups, parametric (Student t-test) or non-parametric (Mann-Whitney U) analysis may be required. One issue relating to these analyses is that, as with genomic studies, the number of lipid variables to be interrogated requires consideration, as this can lead to false assignment of significance. For example, out of 200 lipid variables, 10 will show a difference between any two groups at a significance of  $p = 0.05$  by chance alone. A Bonferroni correction can be used to calculate the  $p$  value required for any individual lipid analyte to be significantly different between groups ( $p = 0.05/200 = 0.00025$ ). However, the correlative nature of lipidomics data makes these calculations inappropriate and likely to miss significant differences; an often used alternative is to perform cross-validation or to validate all differences on an independent cohort.

Multivariate analysis can be used for classification of cases (visualised as a scores plot) and can also provide a measure of the contribution of each lipid variable to the classification model (visualised as a loadings plot). Principal component analysis is the most commonly used method to perform unsupervised analyses, while partial least squares discriminant analysis (PLS-DA) and variants of this provide for supervised analysis where the group identity of the individual samples is used in the model building stage. This approach can also be used for the development of predictive models to classify unknown samples. However, as the number of variables often exceeds the number of samples, caution must be exercised to avoid overfitting the data, such that the resulting model correctly classifies the data set, but has little predictive power for independent samples. Here also, validation is a key issue and must be performed, ideally with independent data sets for all lipidomic analyses.

Another approach to prevent overfitting, while at the same time reducing the complexity of lipid profiles, is to apply recursive feature elimination to develop multivariate models of varying feature size using support vector machine learning. This can provide a ranked list of lipids according

to the frequency of their recurrent incorporation into the generated models and so allows the removal of those variables that do not contribute significantly to the model.

An end goal of many lipidomic studies is the interpretation of the lipidomics data with respect to known biochemical and/or regulatory pathways. Pathway analysis of lipidomic data is still in its infancy. While some tools are available including the KEGG pathway database (<http://www.genome.jp/kegg>), PubChem (<http://pubchem.ncbi.nlm.nih.gov>) and LIPID MAPS (<http://www.lipidmaps.org>), improved bioinformatic tools and methodologies for lipid flux analysis are required to realise the full potential of this approach.

## Lipid Biomarkers in Chronic Disease

A major challenge facing healthcare providers is the growing epidemic of obesity and metabolic syndrome and associated increases in diabetes and heart disease. These complex diseases are influenced by many factors, including genetics, diet and lifestyle. Dyslipidemia is a major feature, usually preceding the clinical onset of disease.

One of the most widely used lipid biomarkers has been cholesterol, which, in the form of total blood cholesterol and/or high density lipoprotein (HDL) cholesterol, has been used in risk calculations for heart disease for over 50 years. Triglycerides are also used clinically for risk assessment of heart disease and diabetes. As we move into the lipidomics era, the potential to accurately and rapidly measure hundreds of individual lipid species provides the opportunity to use more complex lipid profiles as biomarkers of chronic disease. Our laboratory has a focus on the development of lipid profiles to diagnose and assess the risk of chronic disease.

Importantly, many of the risk factors contributing to disease are likely to also influence lipid metabolism and consequently will be reflected in the lipid profile of an individual. The challenge for researchers will be to characterise the profiles that best reflect disease status or risk and translate those into clinically relevant tools.

**Table 2. Lipid-specific mass spectrometry experiments.** Precursor ion scan – PIS; neutral loss scan – NL.

Lipid Group	Parent Ion	MS/MS Experiment, Fragment	Reference
Ceramides	[M+H] <sup>+</sup>	PIS, $m/z$ 264, sphingosine-H <sub>2</sub> O	(7)
Glycolipids	[M+H] <sup>+</sup>	PIS, $m/z$ 264, sphingosine-H <sub>2</sub> O	(7)
Sphingomyelins	[M+H] <sup>+</sup>	PIS, $m/z$ 184, phosphocholine	(7)
Triacylglycerides	[M+NH <sub>4</sub> ] <sup>+</sup>	NL of fatty acids	(16)
Diacylglycerides	[M+NH <sub>4</sub> ] <sup>+</sup>	NL of fatty acids	(16)
Cholesterol esters	[M+NH <sub>4</sub> ] <sup>+</sup>	PIS, $m/z$ 369, cholestane	(17)
Phosphatidic acids	[M-H] <sup>-</sup>	PIS, $m/z$ 153, glycerophosphate-H <sub>2</sub> O	(18)
Phosphatidylcholines	[M+H] <sup>+</sup>	PIS, $m/z$ 184, phosphocholine	(10)
Phosphatidylethanolamines	[M+H] <sup>+</sup>	NL, 141 Da, phosphoethanolamine	(10)
	[M-H] <sup>-</sup>	PIS, $m/z$ 196, dilyso phosphoethanolamine-H <sub>2</sub> O	(18)
Phosphatidylserines	[M+H] <sup>+</sup>	NL, 185 Da, phosphoserine	(18)
	[M-H] <sup>-</sup>	NL, 87 Da, serine-H <sub>2</sub> O	(18)
Phosphatidylglycerols	[M+NH <sub>4</sub> ] <sup>+</sup>	NL, 189 Da, glycerophosphate+NH <sub>4</sub>	(19)
	[M-H] <sup>-</sup>	PIS, $m/z$ 153, glycerophosphate-H <sub>2</sub> O	(18)
Phosphatidylinositols	[M-H] <sup>-</sup>	PIS, $m/z$ 241, inositolphosphate-H <sub>2</sub> O	(18)

## References

1. Fahy, E., Subramaniam, S., Brown, H.A., Glass, C.K., Merrill, A.H., Jr., Murphy, R.C., Raetz, C.R., Russell, D.W., Seyama, Y., Shaw, W., Shimizu, T., Spener, F., van Meer, G., VanNieuwenhze, M.S., White, S.H., Witztum, J.L., and Dennis, E.A. (2005) *J. Lipid Res.* **46**, 839-861
2. Folch, J., Lees, M., and Sloane Stanley, G.H. (1957) *J. Biol. Chem.* **226**, 497-509
3. Yoo, H.H., Son, J., and Kim, D.H. (2006) *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **843**, 327-333
4. Leskinen, H., Suomela, J.P., and Kallio, H. (2007) *Rapid Commun. Mass Spectrom.* **21**, 2361-2373
5. Shaner, R.L., Allegood, J.C., Park, H., Wang, E., Kelly, S., Haynes, C.A., Sullards, M.C., and Merrill, A.H., Jr. (2009) *J. Lipid Res.* **50**, 1692-1707
6. Christie, W.W. (1998) *Lipids* **33**, 343-353
7. Murphy, R.C., Fiedler, J., and Hevko, J. (2001) *Chem. Rev.* **101**, 479-526
8. Han, X., and Gross, R.W. (2003) *J. Lipid Res.* **44**, 1071-1079
9. Hsu, F.F., and Turk, J. (2009) *Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **877**, 2673-2695
10. Pulfer, M., and Murphy, R.C. (2003) *Mass Spectrom. Rev.* **22**, 332-364
11. Gross, R.W., and Han, X. (2009) *Am. J. Physiol. Endocrinol. Metab.* **297**, E297-303
12. Han, X., and Gross, R.W. (2005) *Mass Spectrom. Rev.* **24**, 367-412
13. Yetukuri, L., Katajamaa, M., Medina-Gomez, G., Seppanen-Laakso, T., Vidal-Puig, A., and Oresic, M. (2007) *BMC Syst. Biol.* **1**, 12
14. Jackson, S.N., and Woods, A.S. (2009) *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **877**, 2822-2829
15. Niemela, P.S., Castillo, S., Sysi-Aho, M., and Oresic, M. (2009) *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **877**, 2855-2862
16. Murphy, R.C., James, P.F., McAnoy, A.M., Krank, J., Duchoslav, E., and Barkley, R.M. (2007) *Anal. Biochem.* **366**, 59-70
17. Griffiths, W.J., and Wang, Y. (2009) *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **877**, 2778-2805
18. Cui, Z., and Thomas, M.J. (2009) *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **877**, 2709-2715
19. Meikle, P.J., Duplock, S., Blacklock, D., Whitfield, P.D., Macintosh, G., Hopwood, J.J., and Fuller, M. (2008) *Biochem. J.* **411**, 71-78

