Journal of Developmental Origins of Health and Disease (2015), **6**(2), 65–78.

doi:10.1017/S204017441500001X

REVIEW

Metabolomics in the developmental origins of obesity and its cardiometabolic consequences

M. F. Hivert[1,2†], W. Perng[1†*], S. M. Watkins[3], C. S. Newgard[4], L. C. Kenny[5], B. S. Kristal[6], M. E. Patti[7], E. Isganaitis[7], D. L. DeMeo[8], E. Oken[1,9] and M. W. Gillman[1,9]

[1]*Obesity Prevention Program, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA*
[2]*Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA*
[3] *Metabolon Inc., West Sacramento, CA, USA*
[4]*Nutrition and Metabolism Center, Duke University of Medicine, Durham, NC, USA*
[5]*The Irish Center for Fetal and Neonatal Translational Research, University College Cork, Co. Cork, USA*
[6]*Departments of Neurosurgery, Brigham & Women's Hospital and Harvard Medical School, Boston, MA, USA*
[7]*Research Division, Joslin Diabetes Center, Boston, MA, USA*
[8]*Department of Medicine, Channing Division of Network Medicine and Division of Pulmonary and Critical Care, Brigham and Women's Hospital, Boston, MA, USA*
[9]*Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

In this review, we discuss the potential role of metabolomics to enhance understanding of obesity-related developmental origins of health and disease (DOHaD). We first provide an overview of common techniques and analytical approaches to help interested investigators dive into this relatively novel field. Next, we describe how metabolomics may capture exposures that are notoriously difficult to quantify, and help to further refine phenotypes associated with excess adiposity and related metabolic sequelae over the life course. Together, these data can ultimately help to elucidate mechanisms that underlie fetal metabolic programming. Finally, we review current gaps in knowledge and identify areas where the field of metabolomics is likely to provide insights into mechanisms linked to DOHaD in human populations.

Received 18 July 2014; Revised 2 December 2014; Accepted 23 December 2014; First published online 29 January 2015

Key words: DOHaD, metabolic profiling, metabolomics, obesity

Introduction

Recent technological advances have forged a new 'omics' research era encompassing genomics, epigenomics, transcriptomics, proteomics and metabolomics analyses. As the most downstream constituent of the cascade, metabolomics integrates components of systems biology, chemistry, statistics and informatics to study low-molecular-weight compounds in biological tissues and fluids. Metabolite patterns, also known as 'metabolic signatures',[1] are snapshots of dynamic biochemical activities, providing insight into underlying physiological – or pathophysiological – processes.

In the past decade, metabolomics emerged as a versatile tool with a broad range of applications. In this review, we highlight the potential contribution of metabolomics to epidemiologic studies of developmental origins of health and disease (DOHaD), which historically focused on early life deprivation and later life cardiovascular and metabolic disorders.[2–4] DOHaD has since expanded to encompass a wider scope of early life exposures and a longer list of health outcomes at different life stages. Despite substantial growth in this field, major challenges remain including: (1) objectively characterizing exposures that position individuals on trajectories toward health or disease, (2) accurately characterizing adverse cardiometabolic outcomes and (3) elucidating the underlying mechanisms. Metabolomics may serve as a powerful approach to tackle these challenges in the context of obesity and obesity-related DOHaD research (Fig. 1).

In Section 1, we provide a broad technical background on the metabolomics workflow, including analytical instrumentation and statistical procedures, while highlighting challenges, and sources of variability and their impact on findings. In Section 2, we review the literature and discuss how metabolomics could serve as a means to objectively capture exposures and outcomes relevant to DOHaD. We first discuss how metabolite patterns in the mother and infant could shed light on aberrant physiological changes that lead to later-life disease, while recognizing that maternal metabolism may influence offspring metabolic profiles – either directly through the placenta or indirectly via influences on maternal hormones and/or placental metabolism. We also introduce the idea of using metabolomics to more accurately characterize obesity-related phenotypes throughout the life course. In Section 3, we expose current gaps and propose future directions regarding the use of metabolomics to enhance understanding of mechanisms linking early life environment to later life cardiometabolic disorders.

*Address for correspondence: W. Perng, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, 133 Brookline Avenue, 3rd floor, Boston 02215, USA.
(Email wei.perng@gmail.com)
†Contributed equally as first author

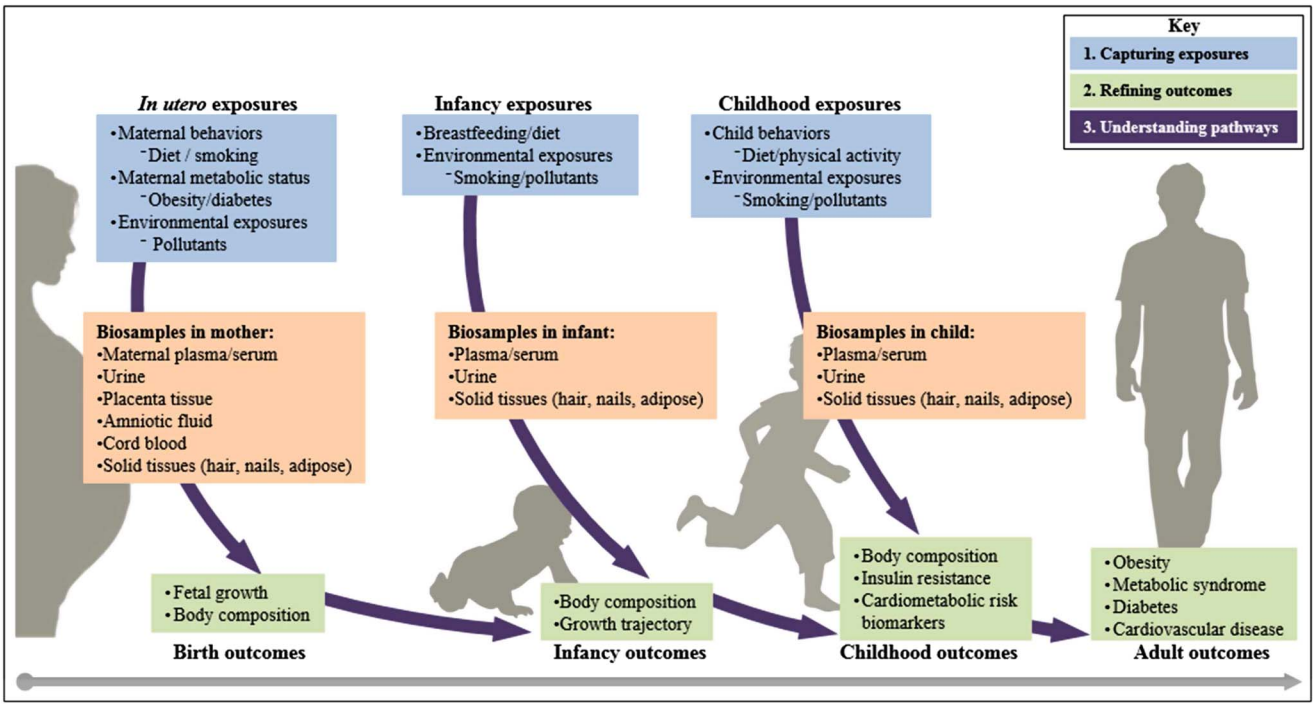

Fig. 1. Areas in which metabolomics could enhance research on developmental origins of obesity and related metabolic disorders

Metabolomics work flow

Metabolomics is the systematic and quantitative analysis of low-molecular-weight intermediates in biological fluids and tissues. Currently, in the Human Metabolome Database,[5] there are over 40,000 documented *metabolites*, which are small molecule intermediates and products of metabolism that include, but are not limited to: amino acids, alcohols, nucleotides, vitamins, fatty acids, antioxidants and organic acids. In this section, we describe the general work flow of metabolomic studies (Fig. 2), discuss frequently used methods and identify areas that merit additional research.

Research question and experimental approach

Metabolomics investigations are either targeted or untargeted. Untargeted studies start with limited *a priori* information regarding the composition of the sample, with the objective of acquiring data on hundreds to thousands of metabolites for the discovery of novel biomarkers. On the other hand, targeted studies are hypothesis-driven and focus on a finite set of metabolites within related biochemical pathways. For example, a researcher conducting a study using an untargeted platform might ask, 'What metabolite patterns are associated with obesity status?' while one pursuing a targeted platform would inquire, 'Are higher levels of branched-chain amino acids associated with obesity status?' We summarize differences between the two approaches in Table 1.

Specimen collection, preparation and storage

The goal of sample collection, preparation and storage is to ensure that the specimen provides a meaningful reflection of the metabolome *in vivo*. This is a critical aspect of any metabolomic study, as failure to collect samples correctly could lead to erroneous findings. Although there are currently no evidence-based guidelines on specimen collection specifically for metabolomics studies, we advise researchers to follow standard operating procedures based on expert consensus for the biosample collection and processing for molecular epidemiological studies.[6–9] Additional considerations for metabolomics sample collection include study design, participant characteristics and tissue type. *Study design* regarding timing of sample collection should take into account the specific questions being asked. For example, fasting specimens provide a steady-state 'snapshot'; however, if the disease process under investigation has metabolic phenotypes which emerge in response to nutrients or exercise (e.g. type 2 diabetes), sampling at defined time points in response to specific interventions may be appropriate. *Participant characteristics* are also important. In pediatric or adolescent populations, one might face challenges of distinguishing between the influence of normal hormonal changes during growth and aberrant metabolite patterns; thus, it may be important to identify pre-pubertal/pubarchal participants from those entering adolescence. Last but not least, the *type of tissue* (e.g. plasma, serum, urine, solid tissue) carries important implications for sample processing (e.g. urea should

https://doi.org/10.1017/S204017441500001X Published online by Cambridge University Press

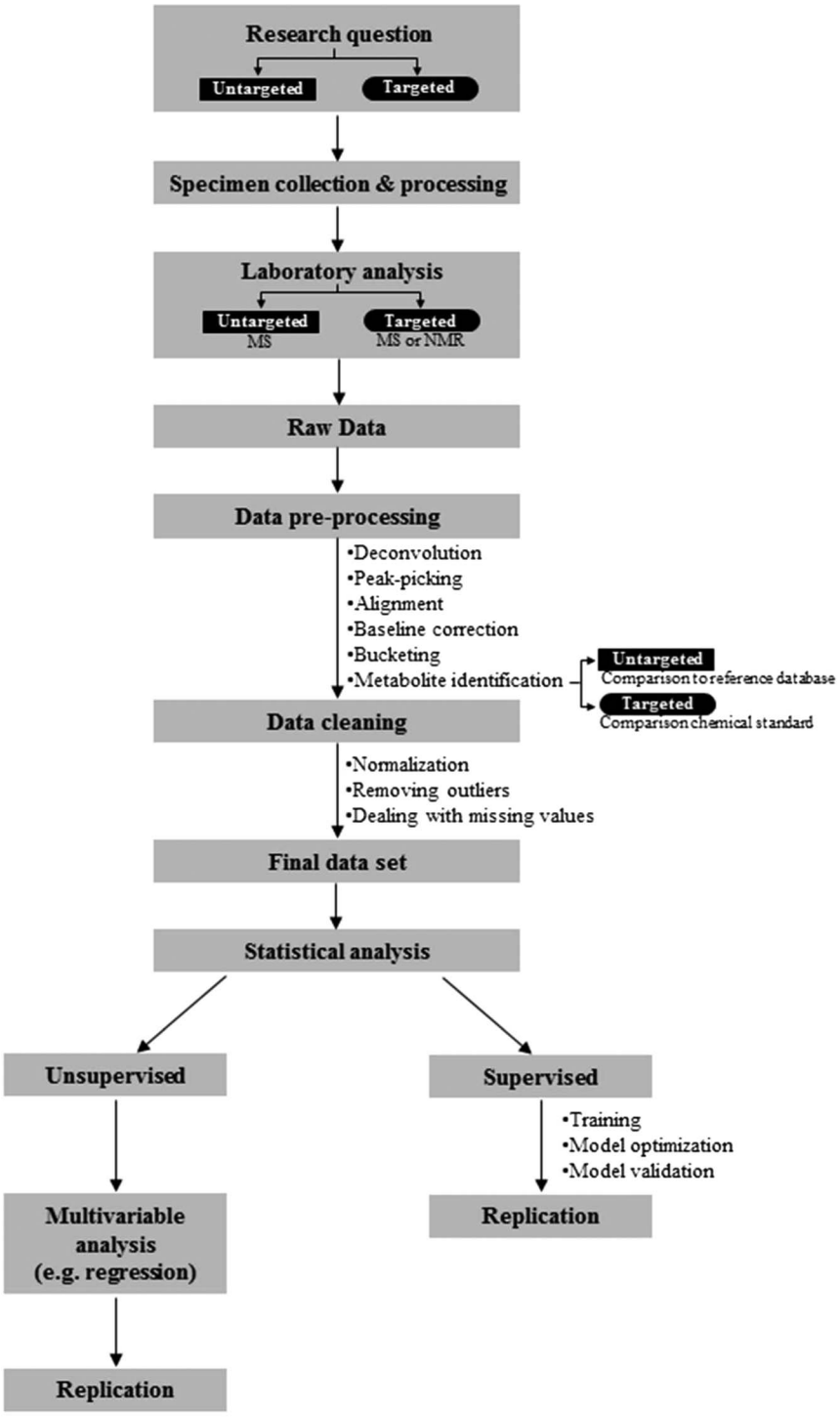

Fig. 2. Metabolomics work flow.

be removed from urine), appropriate anti-coagulants are critical to prepare plasma, and solid tissues must be homogenized before laboratory analysis.

Following collection, tissue or fluid samples should be aliquoted and frozen as soon as possibleat −80°C until time of analysis.[10] Using serum-separator tubes, or via centrifugation before freezing and storage,[10] serum or plasma should be separated from red and white cells. EDTA is a commonly used anti-coagulant, but because it introduces interfering peaks in chromatographic analyses in plasma, experts suggest using

https://doi.org/10.1017/S204017441500001X Published online by Cambridge University Press

Table 1. *Comparison of untargeted* v. *targeted approaches to metabolomics investigations*

	Untargeted	Targeted
Uses	Inductive/hypothesis generating (*discovery*)	Deductive/hypothesis testing
	To obtain global information about numerous biological pathways	To interrogate biochemical pathways based on *a priori* knowledge
	Metabolic profiling; provides a snapshot of metabolome at one point in time	Confirm findings from untargeted studies
Basic characteristics	Semi-quantitative global detection of 100 s to 1000 s of metabolites	Quantification of a small number of metabolites in related metabolic pathways
	Captures multiple biochemical pathways	Captures specific pathways
	Provides relative metabolite concentrations	Provides absolute metabolite concentrations
Advantages	Can quantify a wide range of metabolites	High specificity, precision, and accuracy
	Can lead to discovery of novel biomarkers	Absolute quantification of chemical structures
		High throughput
		Shorter analysis time
Disadvantages	Potential low specificity, precision, and accuracy	Could miss relevant biological pathways
	Metabolite patterns associated with biological condition may require replication using targeted quantitative methods	
	Challenging to identify small metabolite derivatives that do not exist in current chemical libraries	Extensive sample preparation requires inclusion of internal references
Platform options	Mass-spectrometry	Mass-spectrometry or nuclear magnetic resonance spectroscopy

lithium heparin. Avoiding unnecessary freeze-thaw cycles and practicing proper storage techniques will reduce the possibility of metabolite degradation and minimize introduction of systematic error in metabolite profiles.[11–13]

To prepare the specimen for laboratory analysis, the samples should thaw on ice and quenched to minimize enzymatic activity. The metabolites of interest are then extracted from specimen samples. This process is complex and depends on several factors. First, the two most commonly used techniques for metabolomics analyses, mass-spectrometry (MS) and nuclear magnetic resonance (NMR) (as discussed in Section 1.3), operate on different principles and have different requirements for sample processing. NMR exploits the unique nuclear spin of a compound to ascertain its identity and is most effective for fluids,[14] making it necessary to convert solid tissues to liquids before analyses. On the other hand, MS is an analytical chemistry technique that identifies chemicals by measuring the mass-to-charge ratio and is often paired with an ion-separation technique that is applicable to nearly any tissue type. Furthermore, the specimen type itself has important implications for extraction processes; details on extraction processes are available for solid tissue,[15,16] serum and plasma,[16–18] and urine.[16,19] Finally, metabolite class is an important consideration; for example, acetonitrile or perchloric acid should be used to extract polar metabolites, while methanol/chloroform/water is best for lipophilic metabolites.[16]

Analytical instrumentation

There are several technologies available for metabolic phenotyping. The two most frequently used are MS and NMR.[14] Researchers typically couple the MS approach with an ion separation technique, such as gas-liquid chromatography (GC) and/or liquid chromatography (LC). GC separates analytes by vaporizing compounds from the liquid phase into a volatile gas. The gas chromatograph passes the specimen in gaseous form via an inert gas carrier (helium or nitrogen) through a capillary column that contains a stationary liquid adsorbant to promote separation of the molecules as they travel the length of the column. The molecules are retained by the column and elute at different times (known as the 'retention time'). The elution process allows the downstream mass spectrometer to capture and detect individual molecules based on spectral peaks unique to each metabolite's mass-to-charge (*m/z*) ratio. The other commonly used ion-separation technique for MS is LC and the more recent ultra performance liquid chromatography,[20] both of which separate compounds in the soluble phase by passing the specimen as a pressurized liquid through a solid adsorbent-filled column. Comparisons of different MS-based techniques for metabolomics have been reviewed in detail in several scientific articles[21,22] and textbooks.[23]

NMR spectroscopy uses an electromagnetic pulse to induce energy transitions in nuclear spin rates of samples placed in a static magnetic field. When a nucleus is in a magnetic field, its nuclear spin will align in the same (alpha orientation) or opposite (beta orientation) direction of the external magnetic field. Most nuclei exist in the alpha orientation since it requires less energy, so more nuclei are aligned with the external magnetic field than against it. NMR spectroscopy uses electromagnetic radiation to pump the alpha-oriented nuclei into the beta state. When the energy is removed, the nuclei revert or

https://doi.org/10.1017/S204017441500001X Published online by Cambridge University Press

Table 2. *Comparison of mass-spectroscopy (MS) and nuclear magnetic resonance (NMR) spectroscopy platforms for metabolomics*

	MS	NMR spectroscopy
Measured parameter	Mass-to-charge ratio (*m/z*) of metabolite ions separated via	Resonance frequency emitted by the molecule determined by
	Liquid chromatography (LC) or ultra-performance LC (UPLC)	Polarization of the molecule by a constant magnetic field
	Gas chromatography (GC)	Change in energy after electromagnetic pulse
Chemical identification	Ion peaks from chromatograph based on *m/z*, retention time, and ion fragmentation are compared against a standard chemical reference library	Shift in energy levels following application of an electromagnetic pulse in a magnetic field provides structural information on metabolites
Advantages	High sensitivity	High reproducibility
	Ability to detect a wide range of chemicals can be paired with more than one separation technique	No interaction with instrumentation enables multiple analyses on one sample, and re-use later
	Suitable for all specimen types (serum, plasma, urine, solid tissue)	Provides large amount of structural information – especially for larger molecules
		Minimal sample preparation
		Most suitable for biofluids
Disadvantages	Low reproducibility	Low sensitivity, especially for low-molecular-weight compounds
	Physical interaction with instrument can lead to sample contamination	Relatively large amount of sample required
	Can result in overrepresentation of chemicals with multiple structures (increase in false discovery rate)	Large–molecular-weight molecules can obscure spectra for smaller metabolites
		Small detection range (20–24 metabolites in tissue extracts, 30–100 in urine, and 20–30 in plasma or serum)

'relax' back into the alpha state. The fluctuation of the magnetic field associated with the relaxation process is called 'resonance' and can be visualized as peaks in an NMR spectrum to provide structural information about the molecule.[14,24]

The MS and NMR techniques each have their respective strengths and weaknesses,[25,26] as summarized in Table 2. While MS has higher sensitivity and is more versatile in terms of the types of biological specimens it can handle, it tends to capture multiple structures of the same metabolite, making it appear as though more metabolites are involved than in reality. The main advantages to NMR are its ability to acquire quantitative structural information on metabolites, relatively minimal sample preparation, and high reproducibility.[27] The greatest limitation to NMR is insensitivity, as smaller energy transitions from smaller molecules (i.e. sugars, amino acids) are often obscured by those of larger molecules (i.e. phospholipids, triglycerides, lipoproteins).

The choice of MS *v.* NMR depends on the goal of the study. Generally, the ability of MS-based approaches to detect both large and small compounds makes it a better choice for untargeted analyses. MS can be paired with more than one ion separation technique to accommodate the range of metabolites in a biosample. For example, GC is better suited for analysis of small compounds that are thermally stable and volatile (or can be made volatile through derivatization), such as aromatic and organic amino acids, fatty acid derivatives, steroids, flavanoids), whereas LC is best for lipids, peptides, and nucleotides. On the other hand, because targeted studies focus on a set of metabolites from related biochemical pathways, they can typically be carried out with either NMR or a single MS separation technique.

Data pre-processing and cleaning

Metabolite identification

The raw output from MS and NMR platforms contains thousands of signals from detected molecules that require pre-processing and cleaning before conventional statistical analyses. While these steps vary slightly by choice of analytic instrumentation,[26] we outline the general steps in this section.

First, overlapping peaks in an NMR spectrum or a GC/LC chromatogram is separated into signals from individual metabolites ('deconvolution,' now expedited with automated peak filtering technologies[28]). The next steps are to identify peaks that represent actual signals in the NMR or MS chromatogram (*peak picking*)[29] and synchronize the chromatograms[30] or NMR spectra[31] such that each metabolite signal has the same retention time or chemical shift in each sample (*alignment*). The researcher may also choose to correct baseline tilts and drifts in NMR data ('baseline correction,' done automatically or semi-automatically), and define chemical shift bin sizes for NMR spectra (*bucketing*). The final step is to identify compounds in the sample (*metabolite identification*). There are two distinct approaches to metabolite identification, termed *ion-centric* and the *chemocentric*.[32] For untargeted studies, the most common approach is the ion-centric approach, where the

https://doi.org/10.1017/S204017441500001X Published online by Cambridge University Press

researcher compares the signals to a standard reference library, such as the Human Metabolome Database,[33] LipidMaps[34] and MetLin,[35] to identify metabolites. Although this method is less expensive, it runs the risk of misidentifying compounds since these databases do not contain an exhaustive list of all the isotopes of a given metabolite. The chemocentric approach matches the compound to a purified authentic standard. While this approach is markedly more expensive and time consuming, it provides strong reproducibility and validity. In targeted studies where the search is for a defined set of known metabolites, the chemocentric approach is mandatory. Thus, the level of technical expertise at each step influences reliability of the results and their interpretation, a potential contributor to inconsistencies in study findings (discussed in Section 3.1).

Data cleaning

Following pre-processing, data-cleaning steps include normalizing, centering, removing outliers and transforming data to remove noise. During this step, missing values may be imputed. There are two primary sources of missing metabolite data: compound concentrations below the detectable limit of the instrument, and data pre-processing errors such as failure to identify true signals from the background noise. Methods to deal with missingness include replacement with a constant value (e.g. half the minimum detected value), or imputing a range of values using probabilistic Bayesian principal component analysis imputation,[36] Gaussian mixture imputation,[37] and collateral missing value imputation.[38] The laboratory scientist and statistician should discuss the strategy to deal with missing metabolite data as the imputation method has ramifications for statistical analyses.

Statistical analysis

The goal of most metabolomics studies, targeted or untargeted, is to identify metabolites that serve as markers of biological conditions, whether the metabolite pattern precedes that condition (exposure), or whether it transpires as a consequences of the condition (outcome). One approach is to assess relations of each individual metabolite concentration with a binary health condition (e.g. type 2 diabetes) or a continuous biomarker (e.g. fasting insulin levels) using bivariate tests and/or regression analysis. However, these approaches are fraught with the issue of multiple testing since metabolomics studies quantify up to thousands of compounds. Moreover, this concern not easily resolved by corrections for multiple comparisons since metabolites in related pathways will be highly correlated.

One way to deal with the issue of multiple testing is to reduce data dimensionality before conventional analyses. Following data dimensionality reduction, the analyst may examine correlations and use standard regression techniques to assess associations with the exposure or outcome of interest. So-called 'unsupervised' approaches to reduce data dimensionality classify metabolites using the MS *m/z* values or the NMR spectral frequencies as the sole inputs to consolidate compounds into groups based on their interrelations without considering the exposure or outcome. Common examples include principal components analysis (PCA), which generates orthogonal clusters of metabolites that explain the majority of variance in the data;[39] the analysis of variance (ANOVA)-simultaneous component analysis,[40] a combination of ANOVA and PCA methodologies; and cluster analysis techniques, like hierarchical cluster analysis,[41] that group metabolites into natural clusters based on the similarities between each pair of observations. Conventional multivariable analyses may follow these techniques to identify groups of metabolites most strongly associated with the exposure or outcome. Recently, network-based approaches have surfaced as way to intuitively interpret metabolite data using a system of nodes and edges to represent individual molecules and their interactions. Krumsiek *et al.* demonstrated the feasibility using Gaussian graphical models,[42] an undirected probabilistic network modeling technique based on partial pair-wise correlation coefficients, to elucidate the underlying metabolic network structure. A major challenge of unsupervised methods is dealing with unidentified compounds. In this aspect, Gaussian graphical models are useful because they can incorporate metabolomics data with genomic data to infer identities of unknown metabolites.[43,44]

While unsupervised approaches agnostically classify metabolites, the investigator can employ supervised methods when they have an *a priori* hypothesis on the relation between specific metabolites and the biological condition. The goal of supervised methods is to find a model that correctly associates the metabolites with the biological condition of interest. Supervised methods require a training data set to generate a model characterizing the relation between the metabolite patterns and the biological condition. After training, the model is run on an independent data set to assess model validity. Finally, techniques like bootstrapping and cross-validation optimize the model and check for overfitting. Examples of supervised approaches include random forests,[45] a classification method that operates by constructing a multitude of decision trees with the goal of reducing variance, partial least squares-discriminant analysis (PLS-DA; see Fonville *et al.*[46] for details), which relates the independent variable (metabolite data) to the response vector (the biological condition of interest) with a regression model. PLS-DA classifies variables by maximizing the discrimination between predefined sample groups. Another supervised method is orthogonal-PLS,[47] which uses orthogonal signal correction to maximize the explained covariance, providing greater separation between the resulting components. The biggest limitation of supervised approaches is the tendency to overfit the data; thus, testing the model in an independent sample is of paramount importance.

Validation and replication

A crucial aspect of working with high-throughput data is validation and replication of findings. Investigators may

https://doi.org/10.1017/S204017441500001X Published online by Cambridge University Press

improve the internal validity of their results by repeating analyses in the same study sample. A robust validation approach is to first conduct an untargeted study to identify metabolite patterns associated with the biological condition of interest, followed by a targeted study in the same population with the same samples.[48,49] On the other hand, replication of results requires analyses in an independent population – either by splitting the original study sample into 'discovery' and 'replication' groups,[50] or by conducting the study in more than one population,[51] the latter of which is more common. Currently, replication is largely lacking in metabolomics literature – a topic of discussion later in this review.

The role of metabolomics in DOHaD research

The versatility and functionality of metabolomics bestows potential to overcome some classic hurdles of DOHaD research. With assistance from current literature, we discuss ways in which metabolomics may help researchers objectively capture DOHaD exposures, refine cardiometabolic outcomes, and understand the etiological pathways.

Better quantification of DOHaD exposures

Dietary intake

Usual dietary intake. One of the greatest challenges of nutritional assessment is accurately measuring intake. Although food frequency questionnaires, food diaries, and 24-h recalls each have their strengths, they are based on self-report and subject to reporting bias and systematic error.

As the end-products of basic biochemical processes, metabolites provide valuable information on an organism's physiological response to nutrient intake, and could be especially informative for DOHaD research when quantified during key developmental stages like pregnancy and infancy. Although metabolomics studies on nutritional status in pregnant women are rare, a controlled feeding study in non-human primates during the third trimester of pregnancy demonstrated that consumption of a high-fat diet is associated with higher circulating levels of α- and γ-tocopherol,[52] which is in line with observational data from non-pregnant adults that adherence to the healthy eating index[53] is associated with lower γ-tocopherol.[54] Some observational studies in non-pregnant populations examined metabolite profiles associated with dietary patterns[55,56] and intake of specific foods such as fruits and vegetables,[55] garlic, fish and tea.[57] However, these studies were conducted with different designs, populations, and analytic platforms, leading to inconsistent findings. For example, some – but not all – studies found that a Western dietary pattern characterized by high fat and meat intake corresponds with higher levels of circulating trimethylamin-N-oxide (TMAO).[58] In a study of 1003 middle-aged women from the TwinsUK cohort, Menni *et al.* reported that a diet rich in fruits and cruciferous vegetables corresponded with higher serum concentrations of sphingolipids and glycerophospholipids,[55] but despite using the same analytical platform Altamaier *et al.* were unable to replicate these findings in 362 German men.[56] Possible explanations for these discrepancies could be the effect of sex hormones on metabolic profiles, as well as inherent differences – perhaps partially driven by genetics[55] – in the study populations.

There is also evidence from animal studies that the gut microbiome influences diet-related metabolic profile. For example, in rodents, the presence and integrity of the gut microbiota were necessary for diet-induced increases in the metabolite TMAO.[59] Moreover, in a study of twin adult women discordant for obesity, transplant of gut microbes from the obese twin donor into mice led to diet-induced adiposity and elevations in the obesity-related branched chain amino acids (BCAA) metabolite pattern, while transplant of the lean co-twin's microbiota prevented the increase in body mass and obesity-associated phentoypes in the rodents.[60] Preliminary evidence in human adults with metabolic syndrome suggests that microbiota transplants from lean donors improves insulin sensitivity;[61] however, metabolite profiles associated with these interventions in human populations remain to be explored.

Acute response to food intake. The majority of nutritional metabolomics studies are based on blood specimens collected in the fasting state. However, dynamic changes in circulating metabolites following food intake are also relevant and informative. For example, the acute response to an oral glucose load provides information on glucose tolerance, and during pregnancy, it tells of gestational diabetes risk.

Very few metabolomics studies have assessed acute responses to food intake.[62–67] Krug *et al.*[68] administered a range of metabolic challenges to 15 healthy normal weight men over 4 days under controlled conditions for food intake and physical activity. The challenges included a prolonged period of fasting, standardized liquid meal tests, an oral glucose tolerance test (OGTT), an oral lipid tolerance test (OLTT), and exercise testing. Using a targeted metabolomics analysis of lipids and amino acid in blood, urine, exhaled air and breath condensate, the investigators found that anabolic parameters exhibited greater inter-individual variability in the post-prandial states (standardized meals, OGTT or OLTT), whereas catabolic parameters showed the greatest variation after fasting. Thus, while most experts have recommended fasting samples, it is possible that metabolic profiling of non-fasting samples could yield equally important, albeit, different information.

Assessment of dietary interventions. Some investigations in adults have used metabolomics analyses to assess the physiological impact of nutritional interventions and controlled feeding experiments over a defined time period[64] – typically during the course of a few weeks. In DOHaD research, this could be an effective way to quantify dietary modifications in pregnant or lactating women or in the infants themselves, and to assess related physiological changes. For example, in a trial of 1138 European infants, Socha *et al.*[69] found that newborns

https://doi.org/10.1017/S204017441500001X Published online by Cambridge University Press

randomized to receive a high protein formula at ~8 weeks through the 1st year of life had higher serum BCAA by 6 months of age. O'Sullivan *et al.*[70] reported a similar pattern of higher circulating BCAA in formula-fed as compared with breastfed Rhesus monkeys during the 1st weeks of life.[70] A similar BCAA metabolic signature has been associated with insulin resistance and risk of developing type 2 diabetes in adult human studies.[1,51] It currently remains unclear whether elevated BCAAs result from alteration in peripheral metabolism of BCAA,[71] increased proteolysis associated with insulin resistance, or gut-derived factors.[72,60] Nevertheless, these studies not only point toward the utility of metabolomics to evaluate physiological responses to dietary changes, but they also raise the possibility that metabolic profiles predisposing individuals for insulin resistance may stem from early life diet.

In pediatric populations, controlled feeding studies are challenging and thus, rare. In a small study ($n = 24$) of 8-year-old boys, Bertram *et al.*[73] detected differences in serum and urine metabolites following a 7-day nutritional challenge of a diet rich in meat *v.* milk protein. Specifically, the investigators observed lower urinary hippurate levels after the milk-rich diet, and higher urinary excretion of creatine, histidine and urea after the meat-rich diet, but minimal alterations in serum metabolites. These findings shed light on how the type of biosample provides information on different aspects of physiology. For example, tissues with relatively fast turnover rates like plasma, serum and urine provide systemic information regarding ongoing processes. On the other hand storage tissues like hair and nails might reflect long-term dietary intake and accumulation of environmental exposures.[74] Characterizing the diversity and abundance of metabolites in each tissue type is an ongoing effort.

Maternal adiposity and glycemia

Another set of key DOHaD exposures is maternal metabolic status, often assessed using pre-pregnancy body mass index (BMI), gestational weight gain and gestational glucose tolerance. Because body composition assessment in pregnant women is challenging, maternal metabolic phenotyping could reveal metabolic derangements associated with pre-pregnancy obesity or excess gestational weight gain. In a case-control study of 67 hyperglycemic and 50 normoglycemic women, Scholtens *et al.*[75] used an MS-based approach targeting amino acids paired with an untargeted GC-MS panel to interrogate plasma metabolites during the second trimester of pregnancy. Hyperglycemic mothers had higher circulating BCAA than normoglycemic women, which is consistent with evidence from non-pregnant populations.[1]

Smoking

Pre- and postnatal tobacco exposure engenders several adverse health effects. Although circulating cotinine levels are conventionally used to quantify short-term tobacco smoke exposure, metabolomic profiling may provide a functional physiological 'read-out.' In a study of 1241 men and women from the Cooperative Health Research in the Region of Augsburg (KORA) study, Xu *et al.*[76] quantified baseline metabolite profiles in never, former, and current smokers, and assessed change in metabolite profiles in fasting serum samples of participants who stopped smoking during 7 years of follow-up. In addition to detecting alterations in metabolites involved in oxidative stress and atherosclerotic plaque development in current smokers, the investigators found that the metabolic signatures were reversible among smoking-quitters, suggesting that at least in adults, metabolomics biomarkers are sensitive to changes in exposure status. Given the persistent associations of prenatal tobacco exposure with offspring adiposity and cardiometabolic risk,[77] it would be interesting to see whether metabolic differences are also present.

Physical activity

Physical activity is an emerging DOHaD exposure. Metabolomic studies conducted in adults indicate distinct metabolic signatures associated with exercise, mainly driven by an acute rise in lactate and pyruvate.[78–82] Some investigations found evidence of higher BCAA catabolism and consequently, lower levels in serum and higher excretion in urine in response to physical activity.[81,82] In a series of *in vitro* and *in vivo* studies, Roberts *et al.* identified β-Aminoisobutyric acid as an exercise-induced myocyte metabolite that elicits beneficial effects on glycemia in rodent and human adipose tissue.[83] Less is known regarding how longer-term exercise influences metabolic patterns. This is not only due to differences in the type, intensity, and length of training in each study, but also inter-individual variation in response to exercise.[82,84,85]

While current evidence may not yet provide a strong basis to examine metabolite profiles of physical activity as a main exposure in the context of DOHaD, physical activity should be considered as a covariate in metabolomics studies.

Refining outcomes

Although weight and other anthropometric measurements are simple and non-invasive ways to assess cardiometabolic risk, metabolic disturbances vary widely among persons of similar body size. For example, a unique subset of 'metabolically healthy but obese' persons exhibit favorable lipid, cytokine, and inflammatory profiles despite excess adiposity.[86–88] Researchers have also noted the 'metabolically unhealthy normal weight' phenotype characterized by an adverse metabolic profile (i.e. insulin resistance, inflammation, dyslipidemia) despite being normal weight.[89] Metabolic phenotyping could fine-tune definitions of DOHaD risk phenotypes throughout the life course.

Fetal and early life outcomes

DOHaD researchers have traditionally used size at birth, often determined by standardizing birth weight with gestational age- and

https://doi.org/10.1017/S204017441500001X Published online by Cambridge University Press

sex-specific reference data,[90] as an indicator of intrauterine growth. Infants weighing less than the 10th percentile are categorized as small-for-gestational age (SGA), while those who fall above the 90th percentile are considered large-for-gestational age (LGA). Both ends of the spectrum are associated with greater cardiometabolic risk. However, as with the paradigm of the metabolically healthy-obese and metabolically unhealthy-normal-weight phenotypes, fetal growth is a very crude measure of the intrauterine environment. Furthermore, it is unlikely that all SGA and LGA newborns exhibit metabolic risk, and not all appropriate for gestational age (AGA) babies are metabolically healthy. For instance, a baby who was exposed to both maternal smoking and gestational diabetes might fall into the AGA weight range but be far from metabolically healthy.

Cord blood metabolomic profiling could improve assessment of adverse fetal growth outcomes. Although we are not aware of any metabolomics studies of LGA infants, a few case-control studies compared metabolite profiles in umbilical cord blood at delivery between normal weight newborns and infants that were SGA,[91] low birth weight,[92] very low birth weight[93,94] and intra-uterine growth restricted.[95] These studies revealed alterations in amino acid and lipid metabolites, possibly due to impaired nutrient transfer to the fetus during development. In DOHaD investigations, fetal growth is not only assessed as an outcome, but also considered a determinant of future disease risk; it remains to be shown whether metabolic signatures at birth are predictive of later health outcomes.

Life course outcomes: childhood obesity and cardiometabolic status

One of the more well-established areas of metabolomics pertinent to obesity-related DOHaD is characterization of metabolic signatures associated with insulin resistance.[96] In a seminal study that aimed to describe the metabolic signature of an insulin resistant state, BCAAs clearly separated the cases from the controls.[1] Subsequently, researchers reported elevations in BCAAs before development of insulin resistance[97] and type 2 diabetes[51] in non-pregnant adults, suggesting that BCAAs may serve as an upstream indicator of diabetes risk. While there are fewer pediatric metabolomics studies, McCormack *et al.* used a targeted approach to capture amino acids and reported that higher BCAA at baseline predicted worsening of insulin resistance among the 17 pre-pubescent children over 18 months of follow-up.[97] Likewise, we recently found a positive association between a BCAA metabolite pattern and an androgen steroid hormone pattern with obesity during mid-childhood and several markers of cardiometabolic risk, including greater insulin resistance, triglyceride levels, an altered adipocytokine profile, and inflammation.[98] Furthermore, maternal obesity prior to conception was associated with higher BCAA in offspring in mid-childhood, even after accounting for the child's BMI. Such findings support the utility of metabolomics to provide a more nuanced understanding of metabolic derangements that accompany, and perhaps precede, classic cardiometabolic biomarkers.

Perspectives

Current gaps in literature

Lack of replication

One of the major shortcomings of current literature is lack of replication – both in terms of attempts to replicate findings, as well as the consistency in results. Remedying the former will require conscious effort from researchers and planning for replication during study design development. For example, researchers could split the study population into 'discovery' and 'replication' groups, or they could conduct the investigation in more than one cohort in a collaborative consortium. Not so long ago in genetic epidemiology, association studies of complex metabolic traits and diseases were characterized by many reports from small study populations without much replication. In recent years, consortia of several genome-wide association studies made it possible to combine data from multiple cohorts around the world. Metabolomics still has a ways to go before reaching this stage. First, although some cohorts may be using the same commercial platforms, there is no consensus regarding approaches to data cleaning/pre-processing, methods for data reduction, or statistical approach – all of which directly impact consistency of findings. To address this issue, we must identify sources of variability in each step of the metabolomics pipeline (Fig. 2). Second, investigators have to agree on standardized methods for sample collection and laboratory processing, best practices for analytical techniques, and methods to combine data from common or different platforms. While effort has been initiated by the Metabolomics Standards Initiative and the Chemical Analysis Working Group,[99] considerable work is still needed to achieve consensus.

Lack of longitudinal and experimental studies

Another major limitation in the existing literature is the lack of longitudinal investigations, which are crucial to ascertaining temporal relations. One way to clarify temporality is to obtain repeated samples to evaluate how changes in metabolic signatures correspond with hormonal/physiological changes. Among the few examples, metabolite patterns measured at 7-year intervals in non-pregnant adults from the KORA study showed that they were relatively stable over time.[100] In a recent study of 180 healthy pregnant women in China, Luan *et al.*[101] found marked within-person changes in metabolites involved in biopterin, phospholipid, amino acid and fatty acid metabolism across the three trimesters of pregnancy.[94,96–101] The other type of longitudinal design assesses the biological outcome after the exposure. For example, McCormack *et al.* assessed the relation of metabolite patterns measured at baseline with development of insulin resistance in children. However, that study population comprised only 17 participants and the follow-up relatively short (18 months),[97] emphasizing the need for additional prospective investigations during growth and development. Experimental studies are also helpful to provide

https://doi.org/10.1017/S204017441500001X Published online by Cambridge University Press

insight on metabolic adaptations to interventions, such as dietary modifications and exercise. As evidenced in Section 2, these are rare in non-pregnant adults and almost non-existent in pregnant women and children.

Focusing on the strengths of metabolomics

Metabolomics is poised to enhance our understanding of cardiometabolic disease etiology in several ways. First, it provides information not only on how metabolite patterns relate to a biological condition, but also on the interrelations of metabolites in a biosample, and with respect to the exposure(s)/outcome(s) of interest, providing insight on components and interactions of relevant biochemical pathways. Considering that physiological consequences are often a result of alterations in several molecular pathways, these holistic approaches provide a more comprehensive portrait than single biomarkers.

Second, metabolomic analyses can be carried out in nearly any biosample (e.g. blood, urine, serum, hair, nails, saliva, adipose tissue), enabling assessment of how physiological disturbances manifest in tissue specimens. Dyadic bio-samples from mothers and children can be used for metabolic profiling analyses. A few studies have assessed maternal blood and cord blood[91,92,94] to make inference on nutrient exchange and placental function among growth restricted infants. There have also been a few studies assessing metabolite profiles of amniotic fluid and urine of women who developed gestational diabetes, although the findings have not been consistent and the evidence is limited by small sample size.[102,103]

Third, metabolomics can provide insight into DOHaD pathways. Several placental metabolomics studies assessed metabolic phenotypes of hypoxic conditions, including some *in vitro* investigations testing the effect of hypoxia on placental explants.[104–107] These studies pointed out similarities in placental metabolite profiles between hypoxic conditions and pre-eclampsia[106] or intra-uterine growth restriction,[104] implicating their involvement in the pathogenesis of these adverse birth outcomes.

Finally, although the dynamic nature of metabolites is challenging, it represents an opportunity to evaluate the effects of short-term exposures, such as hyperglycemia during the oral glucose tolerance test,[68] where we know that maternal post-load hyperglycemia is associated with adverse fetal outcomes. Exposing specific dyregulated pathways during post-prandial states, or detecting differences in hyperglycemic pregnant women in fasting *v.* post-prandial states will enhance prognosis and management.

In spite of these strengths, it is important to keep limitations of data interpretation in mind. For example, biofluids reflect several concomitant physiological processes and thus preclude the ability to distinguish the precise source of aberrant metabolite patterns. One must exercise caution when speculating on the underlying cause of alterations in metabolism, whether they are due to increased production *v.* lower utilization, dietary intake, the gut microbiome or genetic differences.

Future directions

Metabolomics, in combination with the other omics fields, may help to elucidate the biological processes that underlie the complexity of DOHaD etiology. For example, a recent study used integrated physiology to assess fuel utilization and metabolic flexibility, and metabolomic and lipidomic profiling to generate hypotheses and identify targets for subsequent experimental studies.[108] Similar approaches that integrate metabolomics with genomics, epigenomics, transcriptomics and proteomics, will reveal a more complete picture of DOHaD mechanisms. Metabolites produced by gut microbiota also deserve further investigation based on current evidence regarding its contribution to energy regulation, and because it could be a key mediator between early life factors such as maternal microbiota, mode of delivery, and early infant feeding practices and later-life outcomes.[109]Additionally, genetics concepts such as Mendelian randomization raise the possibility of combining genetic and metabolomics data to generate information on unidentified metabolites[43,44] or to determine causality between metabolite patterns and physiological outcomes.[110,111] Other promising avenues leverage metabolomics for discovery and hypothesis generation, with follow-up studies in functional *in vitro* models or animal studies. For example, after identifying elevations in the furan fatty acid metabolite 3-carboxy-4-methyl-5- propyl-2-furanpropanoic acid (CMPF) in women with getational diabetes, Prentice *et al.* conducted a series of cellular and animal experiments to demonstrate that CMPF could induce β-cell dysfunction, and potentially contribute to progression to type 2 diabetes.[112]

Conclusions

Metabolomics has potential to be a powerful tool for DOHaD research, but variability in approaches to each step of the work flow has yielded inconsistent results that make the literature difficult to interpret. At this point, we can make the following recommendations for DOHaD investigators interested in using metabolomics for their research: (1) clearly define research aims *a priori* (e.g. hypothesis generation *v.* hypothesis testing), (2) plan for replication with adequate sample sizes and (3) when forming the research team, include laboratory, biochemical and statistics/bioinformatics expertise.

As the field of metabolomics continues to grow, advancements should be driven by experimental evidence on best technical and analytical practices. The challenges in the next phase of metabolomics investigations are to efficiently and accurately harvest high-throughput metabolite data, and to integrate them with genomic, transcriptomic and proteomic profiles. In the future, collaborative efforts from multiple cohorts and from consortia of metabolomics data will improve power and generalizability of findings, eventually leading to a better understanding of perturbations that lead to chronic conditions such as type 2 diabetes, obesity and cardiovascular disease.

Acknowledgments

The authors would like to thank Ryan D. Michalek, PhD for his feedback on the manuscript. They are also grateful to Clary Clish, PhD and Andrea Baccarelli, MD PhD for their initial insights on the use of metabolomics for DOHaD research.

Conflicts of Interest

B.S.K. is the inventor on general metabolomics-related IP that has been licensed to Metabolon via Weill Medical College of Cornell University and for which he receives royalty payments via Weill Medical College of Cornell University. He also consults for and has a small equity interest in the company. Metabolon offers biochemical profiling services and is developing molecular diagnostic assays detecting and monitoring disease. Metabolon has no rights or proprietary access to the research results presented and/or new IP generated under these grants/studies. B.S.K.'s interests were reviewed by the Brigham and Women's Hospital and Partners Healthcare in accordance with their institutional policy. Accordingly, upon review, the institution determined that B.S.K.'s financial interest in Metabolon does not create a significant financial conflict of interest (FCOI) with this research. The addition of this statement where appropriate was explicitly requested and approved by BWH.

Funding

EO and WP were supported by the National Institutes of Health (K24 HD069408). WP was additionally supported by the Thomas O. Pyle Fellowship Fund, Harvard Pilgrim Health Care Institute. Grant: US NIH (R37 HD 034568).

References

1. Newgard CB, An J, Bain JR, *et al*. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab*. 2009; 9, 311–326.
2. Barker DJ, Winter PD, Osmond C, Margetts B, Simmonds SJ. Weight in infancy and death from ischaemic heart disease. *Lancet*. 1989; 2, 577–580.
3. Barker DJ, Osmond C. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet*. 1986; 1, 1077–1081.
4. Barker DJ, Gluckman PD, Godfrey KM, *et al*. Fetal nutrition and cardiovascular disease in adult life. *Lancet*. 1993; 341, 938–941.
5. Wishart DS, Jewison T, Guo AC, *et al*. HMDB 3.0 – The Human Metabolome Database in 2013. *Nucleic Acids Res*. 2013; 41 (Database issue) D801–D807.
6. Tuck MK, Chan DW, Chia D, *et al*. Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. *J Proteome Res*. 2009; 8, 113–117.
7. Holland NT, Smith MT, Eskenazi B, Bastaki M. Biological sample collection and processing for molecular epidemiological studies. *Mutat Res*. 2003; 543, 217–234.
8. John MW. ed. Metabolomics methods and protocols. In *Methods in Molecular Biology* (ed. Weckwerth W), 2007; pp. 3–7. Humana Press: Totowa, NJ.
9. Dunn WB, Broadhurst D, Begley P, *et al*. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc*. 2011; 6, 1060–1083.
10. Dietmair S, Timmins NE, Gray PP, Nielsen LK, Kromer JO. Towards quantitative metabolomics of mammalian cells: development of a metabolite extraction protocol. *Anal Biochem*. 2010; 404, 155–164.
11. Teahan O, Gamble S, Holmes E, *et al*. Impact of analytical bias in metabonomic studies of human blood serum and plasma. *Anal Chem*. 2006; 78, 4307–4318.
12. Saude E, Sykes B. Urine stability for metabolomic studies: effects of preparation and storage. *Metabolomics*. 2007; 3, 19–27.
13. Yin P, Peter A, Franken H, *et al*. Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. *Clin Chem*. 2013; 59, 833–845.
14. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev*. 2011; 40, 387–426.
15. Wu H, Southam AD, Hines A, Viant MR. High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Anal Biochem*. 2008; 372, 204–212.
16. Beckonert O, Keun HC, Ebbels TM, *et al*. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc*. 2007; 2, 2692–2703.
17. Want EJ, O'Maille G, Smith CA, *et al*. Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal Chem*. 2006; 78, 743–752.
18. Bruce SJ, Tavazzi I, Parisod V, *et al*. Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. *Anal Chem*. 2009; 81, 3285–3296.
19. Gika HG, Theodoridis G, Extance J, Edge AM, Wilson ID. High temperature-ultra performance liquid chromatography-mass spectrometry for the metabonomic analysis of Zucker rat urine. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2008; 871, 279–287.
20. Wu N, Clausen AM. Fundamental and practical aspects of ultrahigh pressure liquid chromatography for fast separations. *J Sep Sci*. 2007; 30, 1167–1182.
21. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev*. 2007; 26, 51–78.
22. Dunn WB. Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys Biol*. 2008; 5, 011001.
23. Robertson DG and Lindaon J. Metabonomics in Toxicity Assessment. 2005. CRC Press: Boca Raton, FL.
24. Edwards J. Principles of NMR [online]. Retrieved 3 March 2013 from http://www.process-nmr.com/nmr1.htm
25. Lenz EM, Wilson ID. Analytical strategies in metabonomics. *J Proteome Res*. 2007; 6, 443–458.
26. Issaq HJ, Van QN, Waybright TJ, Muschik GM, Veenstra TD. Analytical and statistical approaches to metabolomics research. *J Sep Sci*. 2009; 32, 2183–2199.

27. Dumas ME, Maibaum EC, Teague C, *et al*. Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. *Anal Chem*. 2006; 78, 2199–2208.
28. Scheltema R, Decuypere S, Dujardin J, *et al*. Simple data-reduction method for high-resolution LC-MS data in metabolomics. *Bioanalysis*. 2009; 1, 1551–1557.
29. Katajamaa M, Oresic M. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics*. 2005; 6, 179.
30. Skov T, van den Berg F, Tomasi G, Bro R. Automated alignment of chromatographic data. *J Chemom*. 2006; 20, 484–497.
31. Forshed J, Torgrip RJ, Aberg KM, *et al*. A comparison of methods for alignment of NMR peaks in the context of cluster analysis. *J Pharm Biomed Anal*. 2005; 38, 824–832.
32. Evans AM, Mitchell MW, Dai H and DeHaven C. Categorizing ion – features in liquid chromatography/mass spectrometry metobolomics data. *J Postgenom*. 2012; 2:3.
33. Wishart DS, Knox C, Guo AC, *et al*. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res*. 2009; 37 (Database issue) D603–D610.
34. Fahy E, Subramaniam S, Murphy RC, *et al*. Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res*. 2009; 50(Suppl.), S9–S14.
35. Smith CA, O'Maille G, Want EJ, *et al*. METLIN: a metabolite mass spectral database. *Ther Drug Monit*. 2005; 27, 747–751.
36. Oba S, Sato MA, Takemasa I, *et al*. A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics*. 2003; 19, 2088–2096.
37. Ouyang M, Welsh WJ, Georgopoulos P. Gaussian mixture clustering and imputation of microarray data. *Bioinformatics*. 2004; 20, 917–923.
38. Sehgal MS, Gondal I, Dooley LS. Collateral missing value imputation: a new robust missing value estimation algorithm for microarray data. *Bioinformatics*. 2005; 21, 2417–2423.
39. Jolliffe IT. *Principal Component Analysis*. 1986. Springer-Verlag: New York.
40. Smilde AK, Jansen JJ, Hoefsloot HC, *et al*. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics*. 2005; 21, 3043–3048.
41. Beckonert O, Bollard ME, Ebbels T, *et al*. NMR-based metabonomic toxicity classification: hierarchical cluster analysis and k-nearest-neighbour approaches. *Analytica Chimica Acta*. 2003; 490, 3–15.
42. Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst Biol*. 2011; 5, 21.
43. Krumsiek J, Suhre K, Evans AM, *et al*. Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. *PLoS Genet*. 2012; 8, e1003005.
44. Shin SY, Fauman EB, Petersen AK, *et al*. An atlas of genetic influences on human blood metabolites. *Nat. Genet*. 2014; 46, 543–550.
45. Truong Y, Lin X, Beecher C. Learning a complex metabolomic dataset using random forests and support vector machines. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004; Seattle, WA, USA.
46. Fonville JM, Richards SE, Barton RH, *et al*. The evolution of partial least squares models and related chemometric approaches in metabonomics and metabolic phenotyping. *J Chemom*. 2010; 24, 636–649.
47. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). *J Chemom*. 2002; 16, 119–128.
48. Shi H, Vigneau-Callahan KE, Shestopalov AI, *et al*. Characterization of diet-dependent metabolic serotypes: primary validation of male and female serotypes in independent cohorts of rats. *J Nutr*. 2002; 132, 1039–1046.
49. Shi H, Vigneau-Callahan KE, Shestopalov AI, *et al*. Characterization of diet-dependent metabolic serotypes: proof of principle in female and male rats. *J Nutr*. 2002; 132, 1031–1038.
50. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*. 2006; 38, 209–213.
51. Wang TJ, Larson MG, Vasan RS, *et al*. Metabolite profiles and the risk of developing diabetes. *Nat Med*. 2011; 17, 448–453.
52. Cox J, Williams S, Grove K, Lane RH, Aagaard-Tillery KM. A maternal high-fat diet is accompanied by alterations in the fetal primate metabolome. *Am J Obstet Gynecol*. 2009; 201, 281. e281–281.e289.
53. Kennedy ET, Ohls J, Carlson S, Fleming K. The Healthy Eating Index: design and applications. *J Am Diet Assoc*. 1995; 95, 1103–1108.
54. Guertin KA, Moore SC, Sampson JN, *et al*. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *Nat Genet*. 2014; 100, 208–217.
55. Menni C, Zhai G, Macgregor A, *et al*. Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics*. 2013; 9, 506–514.
56. Altmaier E, Kastenmuller G, Romisch-Margl W, *et al*. Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. *Eur J Epidemiol*. 2011; 26, 145–156.
57. Floegel A, von Ruesten A, Drogan D, *et al*. Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. *Eur J Clin Nutr*. 2013; 67, 1100–1108.
58. Steffen LM, Zheng Y, Steffen BT. Metabolomic biomarkers reflect usual dietary pattern: a review. *Curr Nutr Rep*. 2014; 3, 62–68.
59. Wang Z, Klipfell E, Bennett BJ, *et al*. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011; 472, 57–63.
60. Ridaura VK, Faith JJ, Rey FE, *et al*. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science*. 2013; 341, 1241214.
61. Vrieze A, Van Nood E, Holleman F, *et al*. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology*. 2012; 143, 913–916, e917.
62. Llorach R, Urpi-Sarda M, Jauregui O, Monagas M, Andres-Lacueva C. An LC-MS-based metabolomics approach for exploring urinary metabolome modifications after cocoa consumption. *J Proteome Res*. 2009; 8, 5060–5068.
63. van Velzen EJ, Westerhuis JA, van Duynhoven JP, *et al*. Phenotyping tea consumers by nutrikinetic analysis of polyphenolic end-metabolites. *J Proteome Res*. 2009; 8, 3317–3330.

64. Johansson-Persson A, Barri T, Ulmius M, Onning G, Dragsted LO. LC-QTOF/MS metabolomic profiles in human plasma after a 5-week high dietary fiber intake. *Anal Bioanal Chem*. 2013; 405, 4799–4809.
65. Gurdeniz G, Rago D, Bendsen NT, *et al*. Effect of trans fatty acid intake on LC-MS and NMR plasma profiles. *PLoS One*. 2013; 8, e69589.
66. Schmidt MD, Dwyer T, Magnussen CG, Venn AJ. Predictive associations between alternative measures of childhood adiposity and adult cardio-metabolic health. *Int J Obes (Lond)*. 2011; 35, 38–45.
67. Bondia-Pons I, Nordlund E, Mattila I, *et al*. Postprandial differences in the plasma metabolome of healthy Finnish subjects after intake of a sourdough fermented endosperm rye bread versus white wheat bread. *Nutr J*. 2011; 10, 116.
68. Krug S, Kastenmuller G, Stuckler F, *et al*. The dynamic range of the human metabolome revealed by challenges. *FASEB J*. 2012; 26, 2607–2619.
69. Socha P, Grote V, Gruszfeld D, *et al*. Milk protein intake, the metabolic-endocrine response, and growth in infancy: data from a randomized clinical trial. *Am J Clin Nutr*. 2011; 94(6 Suppl.), 1776s–1784s.
70. O'Sullivan A, He X, McNiven EM, *et al*. Early diet impacts infant rhesus gut microbiome, immunity, and metabolism. *J Proteome Res*. 2013; 12, 2833–2845.
71. Herman MA, She P, Peroni OD, Lynch CJ, Kahn BB. Adipose tissue branched chain amino acid (BCAA) metabolism modulates circulating BCAA levels. *J Biol Chem*. 2010; 285, 11348–11356.
72. Lu J, Xie G, Jia W, Jia W. Insulin resistance and the metabolism of branched-chain amino acids. *Front Med*. 2013; 7, 53–59.
73. Bertram HC, Hoppe C, Petersen BO, *et al*. An NMR-based metabonomic investigation on effects of milk and meat protein diets given to 8-year-old boys. *Br J Nutr*. 2007; 97, 758–763.
74. Scheepers PT. The use of biomarkers for improved retrospective exposure assessment in epidemiological studies: summary of an ECETOC workshop. *Biomarkers*. 2008; 13, 734–748.
75. Scholtens DM, Muehlbauer MJ, Daya NR, *et al*. Metabolomics reveals broad-scale metabolic perturbations in hyperglycemic mothers during pregnancy. *Diabetes Care*. 2014; 37, 158–166.
76. Xu T, Holzapfel C, Dong X, *et al*. Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med*. 2013; 11, 60.
77. Oken E, Levitan EB, Gillman MW. Maternal smoking during pregnancy and child overweight: systematic review and meta-analysis. *Int J Obes (Lond)*. 2008; 32, 201–210.
78. Enea C, Seguin F, Petitpas-Mulliez J, *et al*. (1)H NMR-based metabolomics approach for exploring urinary metabolome modifications after acute and chronic physical exercise. *Anal Bioanal Chem*. 2010; 396, 1167–1176.
79. Lewis GD, Farrell L, Wood MJ, *et al*. Metabolic signatures of exercise in human plasma. *Sci Transl Med*. 2010; 2, 33ra37.
80. Netzer M, Weinberger KM, Handler M, *et al*. Profiling the human response to physical exercise: a computational strategy for the identification and kinetic analysis of metabolic biomarkers. *J Clin Bioinforma*. 2011; 1, 34.
81. Pechlivanis A, Kostidis S, Saraslanidis P, *et al*. (1)H NMR-based metabonomic investigation of the effect of two different exercise sessions on the metabolic fingerprint of human urine. *J Proteome Res*. 2010; 9, 6405–6416.
82. Pechlivanis A, Kostidis S, Saraslanidis P, *et al*. 1H NMR study on the short- and long-term impact of two training programs of sprint running on the metabolic fingerprint of human serum. *J Proteome Res*. 2013; 12, 470–480.
83. Roberts LD, Bostrom P, O'Sullivan JF, *et al*. β-Aminoisobutyric acid induces browning of white fat and hepatic β-oxidation and is inversely correlated with cardiometabolic risk factors. *Cell Metab*. 2014; 19, 96–108.
84. Huffman KM, Slentz CA, Bateman LA, *et al*. Exercise-induced changes in metabolic intermediates, hormones, and inflammatory markers associated with improvements in insulin sensitivity. *Diabetes Care*. 2011; 34, 174–176.
85. Yan B, AJ, Wang G, *et al*. Metabolomic investigation into variation of endogenous metabolites in professional athletes subject to strength-endurance training. *J Appl Physiol (1985)*. 2009; 106, 531–538.
86. Brochu M, Tchernof A, Dionne IJ, *et al*. What are the physical characteristics associated with a normal metabolic profile despite a high level of obesity in postmenopausal women? *J Clin Endocrinol Metab*. 2001; 86, 1020–1025.
87. Karelis AD. Metabolically healthy but obese individuals. *Lancet*. 2008; 372, 1281–1283.
88. Karelis AD, Faraj M, Bastard JP, *et al*. The metabolically healthy but obese individual presents a favorable inflammation profile. *J Clin Endocrinol Metab*. 2005; 90, 4145–4150.
89. Thomas EL, Parkinson JR, Frost GS, *et al*. The missing risk: MRI and MRS phenotyping of abdominal adiposity and ectopic fat. *Obesity (Silver Spring)*. 2012; 20, 76–87.
90. Oken E, Kleinman KP, Rich-Edwards J, Gillman MW. A nearly continuous measure of birth weight for gestational age using a United States national reference. *BMC Pediatr*. 2003; 3, 6.
91. Horgan RP, Broadhurst DI, Walsh SK, *et al*. Metabolic profiling uncovers a phenotypic signature of small for gestational age in early pregnancy. *J Proteome Res*. 2011; 10, 3660–3673.
92. Ivorra C, Garcia-Vicent C, Chaves FJ, *et al*. Metabolomic profiling in blood from umbilical cords of low birth weight newborns. *J Transl Med*. 2012; 10, 142.
93. Alexandre-Gouabau MC, Courant F, Moyon T, *et al*. Maternal and cord blood LC-HRMS metabolomics reveal alterations in energy and polyamine metabolism, and oxidative stress in very-low birth weight infants. *J Proteome Res*. 2013; 12, 2764–2778.
94. Tea I, Le Gall G, Kuster A, *et al*. 1H-NMR-based metabolic profiling of maternal and umbilical cord blood indicates altered materno-foetal nutrient exchange in preterm infants. *PLoS One*. 2012; 7, e29947.
95. Favretto D, Cosmi E, Ragazzi E, *et al*. Cord blood metabolomic profiling in intrauterine growth restriction. *Anal Bioanal Chem*. 2012; 402, 1109–1121.
96. Morris C, O'Grada C, Ryan M, *et al*. The relationship between BMI and metabolomic profiles: a focus on amino acids. *Proc Nutr Soc*. 2012; 71, 634–638.
97. McCormack SE, Shaham O, McCarthy MA, *et al*. Circulating branched-chain amino acid concentrations are associated with obesity and future insulin resistance in children and adolescents. *Pediatr Obes*. 2013; 8, 52–61.
98. Perng WGM, Fleisch AF, Michalek RD, *et al*. Metabolomic profiles of childhood obesity. Early Nutrition Conference 2014.

99. Sumner LW, Amberg A, Barrett D, *et al*. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007; 3, 211–221.
100. Yousri NA, Kastenmuller G, Gieger C, *et al*. Long term conservation of human metabolic phenotypes and link to heritability. *Metabolomics*. 2014; 10, 1005–1017.
101. Luan H, Meng N, Liu P, *et al*. Pregnancy-induced metabolic phenotype variations in maternal plasma. *J Proteome Res*. 2014; 13, 1527–1536.
102. Graca G, Goodfellow BJ, Barros AS, *et al*. UPLC-MS metabolic profiling of second trimester amniotic fluid and maternal urine and comparison with NMR spectral profiling for the identification of pregnancy disorder biomarkers. *Mol Biosyst*. 2012; 8, 1243–1254.
103. Graca G, Duarte IF, Barros AS, *et al*. Impact of prenatal disorders on the metabolic profile of second trimester amniotic fluid: a nuclear magnetic resonance metabonomic study. *J Proteome Res*. 2010; 9, 6016–6024.
104. Horgan RP, Broadhurst DI, Dunn WB, *et al*. Changes in the metabolic footprint of placental explant-conditioned medium cultured in different oxygen tensions from placentas of small for gestational age and normal pregnancies. *Placenta*. 2010; 31, 893–901.
105. Heazell AE, Brown M, Dunn WB, *et al*. Analysis of the metabolic footprint and tissue metabolome of placental villous explants cultured at different oxygen tensions reveals novel redox biomarkers. *Placenta*. 2008; 29, 691–698.
106. Dunn WB, Brown M, Worton SA, *et al*. Changes in the metabolic footprint of placental explant-conditioned culture medium identifies metabolic disturbances related to hypoxia and pre-eclampsia. *Placenta*. 2009; 30, 974–980.
107. Tissot van Patot MC, Murray AJ, Beckey V, *et al*. Human placental metabolic adaptation to chronic hypoxia, high altitude: hypoxic preconditioning. *Am J Physiol Regul Integr Comp Physiol*. 2010; 298, R166–R172.
108. Kurland IJ, Accili D, Burant C, *et al*. Application of combined omics platforms to accelerate biomedical discovery in diabesity. *Ann NY Acad Sci*. 2013; 1287, 1–16.
109. Putignani L, Del Chierico F, Petrucca A, Vernocchi P, Dallapiccola B. The human gut microbiota: a dynamic interplay with the host from birth to senescence settled during childhood. *Pediatr Res*. 2014; 76, 2–10.
110. Wurtz P, Kangas AJ, Soininen P, *et al*. Lipoprotein subclass profiling reveals pleiotropy in the genetic variants of lipid risk factors for coronary heart disease: a note on Mendelian randomization studies. *J Am Coll Cardiol*. 2013; 62, 1906–1908.
111. Timpson NJ, Nordestgaard BG, Harbord RM, *et al*. C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. *Int J Obes (Lond)*. 2011; 35, 300–308.
112. Prentice KJ, Luu L, Allister EM, *et al*. The furan fatty acid metabolite CMPF is elevated in diabetes and induces beta cell dysfunction. *Cell Metab*. 2014; 19, 653–666.

https://doi.org/10.1017/S204017441500001X Published online by Cambridge University Press