



Conference on 'Translating nutrition: integrating research, practice and policy' Symposium 2: Intervention study design and personalised nutrition

Inter-individual differences in response to dietary intervention: integrating omics platforms towards personalised dietary recommendations

Johanna W. Lampe¹, Sandi L. Navarro¹, Meredith A. J. Hullar¹ and Ali Shojaie²

¹Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

²Department of Biostatistics, University of Washington, Seattle, WA, USA

Technologic advances now make it possible to collect large amounts of genetic, epigenetic, metabolomic and gut microbiome data. These data have the potential to transform approaches towards nutrition counselling by allowing us to recognise and embrace the metabolic, physiologic and genetic differences among individuals. The ultimate goal is to be able to integrate these multi-dimensional data so as to characterise the health status and disease risk of an individual and to provide personalised recommendations to maximise health. To this end, accurate and predictive systems-based measures of health are needed that incorporate molecular signatures of genes, transcripts, proteins, metabolites and microbes. Although we are making progress within each of these omics arenas, we have yet to integrate effectively multiple sources of biologic data so as to provide comprehensive phenotypic profiles. Observational studies have provided some insights into associative interactions between genetic or phenotypic variation and diet and their impact on health; however, very few human experimental studies have addressed these relationships. Dietary interventions that test prescribed diets in well-characterised study populations and that monitor system-wide responses (ideally using several omics platforms) are needed to make correlation–causation connections and to characterise phenotypes under controlled conditions. Given the growth in our knowledge, there is the potential to develop personalised dietary recommendations. However, developing these recommendations assumes that an improved understanding of the phenotypic complexities of individuals and their responses to the complexities of their diets will lead to a sustainable, effective approach to promote health and prevent disease – therein lies our challenge.

Human feeding study: Metabolomics: Gut microbiome: Data integration

Technologic advances now make it possible to collect large amounts of genetic, epigenetic, proteomic, metabolomic and gut microbiome data. Many of the applications of this multi-dimensional data have been in the areas of disease detection, prognosis and treatment. However, such approaches may also lend themselves towards characterising healthy phenotypes and more effectively informing dietary recommendations for maintaining or improving the health of individuals on a personal level. Omics data have the potential to transform our approach towards nutrition

counselling by allowing us to recognise and embrace the metabolic, physiologic and genetic differences among individuals. The ultimate goal would be to integrate these multi-dimensional data so as to characterise the health status and disease risk of an individual and to provide personalised dietary recommendations to maximise health. To this end, accurate and predictive system-based measures of health are needed that incorporate molecular signatures of genes, transcripts, proteins, metabolites and microbes.

Abbreviation: GST, glutathione S-transferase.

Corresponding author: Dr J. W. Lampe, fax +1 206 667 7850, email jlampe@fhcrc.org

Nutrition, as a science, has a long tradition of determining the nutrient requirements of heterogeneous populations eating a wide variety of diets and of providing dietary recommendations for health. This has typically involved simplifying the inherent complexity into manageable recommendations in the form of dietary guidance for the purpose of preventing disease in a population. Despite the application of biostatistical approaches with the goal to be as inclusive of the population as possible, there are limitations due to assumptions that metabolic organisational structure is uniform among individuals and that direct cause–effect relationships exist. In reality, the large number of functional redundancies and adaptive mechanisms that provide for homeostasis⁽¹⁾ make evaluating the complexities and nuances challenging.

The concept of a ‘nutritional phenotype’, i.e. an integrated set of genetic, proteomic, metabolomic, functional and behavioural factors that, when measured, could provide the basis for assessment of human nutritional status, was introduced several years ago by Ziesel *et al.*⁽²⁾ It was proposed as a way to integrate the effects of diet on disease/wellness and provide a quantitative indication of the paths by which genes and environment exert their effects on health⁽¹⁾. The concept provides a good base from which to begin to establish approaches to personalised dietary recommendations; however, several questions need to be addressed. These include, but are not necessarily limited to: What data will we need on an individual in order to personalise dietary recommendations? How can we use controlled feeding studies and other dietary interventions to generate a nutritional phenotypic framework? How can we most effectively integrate omics data so as to be able to apply them towards personalised nutrition?

What data will we need on an individual in order to personalise dietary recommendations?

Numerous factors contribute to variation in nutritional requirements and responses to diet, including sex, stage of life cycle, disease, physical activity level, genetic background, gut microbial community and environmental exposures. Several of these are already considered in the construction of personalised nutritional recommendations; for example, sex, age, adiposity and activity level are routinely used in determining nutrient requirements in healthy individuals and understanding the contributions of disease state to nutritional requirements is a hallmark of therapeutic nutrition. To date, the more complex factors such as genomics, host microbial community structure and environmental exposures are often not included in the equation.

Genetic polymorphisms are well-recognised sources of variation in human response to some aspects of diet, including taste preference, food tolerance, nutrient absorption, transport and metabolism, and effects at target tissues⁽³⁾. Typically, in past studies, one particular genetic variant has been considered in relation to intake of one particular nutrient. For example, two polymorphisms in the *MTHFR* gene (C677T and A1298C) are associated with reduced methylenetetrahydrofolate reductase activity and

higher homocysteine concentrations⁽³⁾. Carriers of these polymorphisms are at higher risk of CVD; thus sufficient intake of folate is particularly important. Other examples include iron overload and haemochromatosis, copper malabsorption and Menkes disease, and glucose-6-phosphate dehydrogenase and consumption of fava beans, high in pro-oxidant glycosides (favism; reviewed in⁽³⁾). Further, genomics may contribute to phenotypic differences in health behaviour and modify response to interventions designed to change health behaviours⁽⁴⁾.

Several genome-wide association studies have evaluated the association between multiple SNP and metabolomics profiles. In a sample of 284 men, Gieger *et al.*⁽⁵⁾ integrated genome-wide association study data with serum metabolomics-based quantitation of 363 metabolites. They reported associations of frequent SNP with differences in the metabolic homeostasis, explaining up to 12% of the observed variance. Using ratios of certain metabolite concentrations as a proxy for enzymatic activity, up to 28% of the variance can be explained (P -values 10^{-16} – 10^{-21}). Four variants in genes coding for enzymes (FADS1, LIPC, SCAD and MCAD) were identified where a corresponding metabolic phenotype (metabotype) clearly matched the biochemical pathways in which these enzymes are active.

More recently, Suhre *et al.*⁽⁶⁾ conducted an analysis of genotype-dependent metabolic phenotypes using a genome-wide association study with non-targeted metabolomics in a sample of 1768 individuals. They identified thirty-seven genetic loci associated with blood metabolite concentrations, of which twenty-five showed effect sizes that accounted for 10–60% difference in metabolite levels per allele copy. These results provided functional insights into disease-related associations that have been reported in previous studies, including those for cardiovascular and renal disorders, type 2 diabetes, cancer, gout, venous thromboembolism and Crohn’s disease.

The human gut microbial community also shapes host exposure to dietary constituents by modulating absorption, storage and energy harvest from the diet. It is a large, complex ecosystem, with the number of different species of bacteria estimated to range from 300 and 1000 and the majority of the species diversity distributed between the phyla Firmicutes and Bacteroidetes^(7,8). There is high inter-individual variation in the composition of communities, mostly at the species level⁽⁹⁾, whereas the distribution of bacterial functional genes is less varied. This functional redundancy is a hallmark of a stable symbiosis in which many different species carry out the same functional role.

Recent studies suggest that individuals can be clustered into distinct groups based on their gut microbiome composition and functional metabolism⁽¹⁰⁾. The underlying metabolism of the dominant bacteria that define these groups is the degradation of plant polymers (e.g. dietary fibre) via different metabolic pathways; long-term dietary habits have been associated with these groupings⁽¹⁰⁾. Through the metabolism of dietary constituents, the gut microbiome can influence the magnitude and flux of metabolites to which the host is exposed and some of the variations in what have been identified as genotype-dependent metabolic phenotypes actually may be due to the composition and activity of the gut microbiome^(11,12).



Indeed, of the genotype-dependent metabolic phenotypes identified by Suhre *et al.*⁽⁶⁾, an altered microbiome has been associated with CVD⁽¹³⁾, type 2 diabetes⁽¹⁴⁾, some cancers^(15,16) and Crohn's disease⁽¹⁷⁾. However, the relationships between the gut microbiome, diet and metabolic phenotypes need to be addressed in rigorous experimental settings using approaches that integrate metabolomics, host genomics and the gut microbiome.

How can we use controlled feeding studies and other dietary interventions to develop phenotype profiles?

Controlled feeding studies in healthy human subjects have been used for over a century to establish the quantitative requirements and confirm essentiality of nutrients in human subjects. Typically, these studies had small sample sizes, were intensively controlled, and often focused on restriction and re-feeding of specific nutrients or nutrient sources. They were used to evaluate the acute effects of food deprivation, show experimentally the effects of dietary restrictions on development of deficiency diseases, establish specific amino acid requirements and describe vitamin metabolism⁽¹⁸⁾. Consequently, they were crucial in determining recommended daily dietary allowances. Controlled interventions and defined background diets have also been useful for testing response to varying doses of a dietary constituent⁽¹⁹⁾ and for testing and monitoring biomarkers of disease susceptibility and dietary exposure⁽²⁰⁾. More recently, dietary interventions have been used to test the effects of particular dietary patterns⁽²¹⁾ and to test genotype–phenotype interactions⁽²²⁾.

Controlled feeding studies, particularly with randomised crossover designs where each person serves as their own control, are a useful venue in which to test genotype–diet interactions as well as genotype–phenotype associations. In the latter case, the relationship between genotype and phenotype can sometimes be better characterised on the background of the same dietary exposures (i.e. a controlled diet)⁽²²⁾. Participant screening protocols for recruitment into controlled feeding studies also can be set up to enrich *a priori* for particular genotypes or phenotypes so as to provide more equal distributions of sample sizes in subgroups, particularly if the prevalence of a particular variant is low, and to increase statistical power to compare these subgroups.

Controlled feeding studies also provide a useful approach in which to characterise host–gut microbial interactions and to determine gut microbial community response to diet. In the context of controlled dietary interventions, gut bacterial community composition has been shown to differ significantly when participants consume different diets^(23,24), although the overall response of the gut bacterial community is often unique for each individual^(24,25). Most studies have tested effects of fermentable complex carbohydrates (e.g. dietary fibres, resistant starch; Table 1). Network analysis of the gut microbial community reveals niche specialisation based on a metabolic interconnection between different bacteria that are often specialised in one enzymatic transformation in the pathway of dietary metabolism^(10,26–28). The type of carbohydrate

ingested often influences the prevalence of certain groups of gut bacteria and the subsequent composition of the microbial metabolic end products to which the host is exposed (e.g. SCFA; Table 1). Differences in gut microbial metabolism of various phytochemicals also contribute to gut bacterial metabolic phenotypes that influence dietary exposures⁽²⁹⁾. Being able to test for the effects of these phenotypes in the context of nutrition interventions is important, since some subgroups may be more responsive to the intervention than others. For example, Niculescu *et al.*⁽³⁰⁾ reported differential lymphocyte gene expression by bacterial metabolic phenotype in postmenopausal women receiving an isoflavone supplement; a greater increase in oestrogen-responsive genes was observed in women who carried the bacteria capable of converting the soya isoflavone daidzein to equol.

'Omics' – transcriptomics, proteomics and metabolomics – approaches have been hypothesised to revolutionise our understanding of the interactions of the various systems that are often studied in isolation and have the potential to revolutionise many aspects of our study of nutrition and health promotion. Despite the excitement, at this stage, the technologies still require rigorous evaluation and validation. Controlled feeding studies are a useful approach in which to validate and test the robustness of these omics approaches with the goal of ultimately being able to use them to evaluate the effects of totality of diet on totality of response in human subjects. In addition, they provide important details on the behaviour of proteins, transcripts and metabolites under controlled conditions.

Several studies have used the construct of controlled feeding interventions to test effects of diet on omics measures (Table 2). The majority of these have utilised metabolomics to characterise response to phytochemical-containing foods (fruits, vegetables, tea, nuts) compared with a control in healthy individuals. Many of the metabolites identified typically correspond to dietary biomarkers of the intervention foods consumed (e.g. proline betaine after consumption of citrus fruits). Although many studies also yield a handful of endogenous metabolites that differ in abundance between the interventions, these compounds are often generally reported as differences in metabolite profiles owing to a lack of adequate pathway analysis tools. Thus, it is often unclear whether differences in metabolite profiles are indicative of perturbations in specific pathways or molecular targets in response to the dietary intervention, or are unrelated compounds identified by chance. Some investigators have explored pathways manually. For example, Solanky *et al.*⁽³¹⁾ found that soya consumption was associated with osmolyte fluctuations and differences in energy metabolism. Work in our laboratory (DH May, SL Navarro, I Ruczinski *et al.*, unpublished results) suggests potential differences in energy utilisation from glucose to fat between a diet devoid of fruits and vegetables compared with a diet high in crucifers, citrus and soya. These examples provide provocative views of other mechanisms through which plant foods may promote health; however, even with manual analyses, the interpretation is still broad, speculative and incomplete.

Other investigations have employed alternative omics technologies to study response to diet and have evaluated

Table 1. Summary of human dietary randomised crossover intervention studies of response of the gut microbiome to diet

Dietary intervention, reference and year	Sample size/population	Dietary intervention/dose of food agent	Treatment duration	Platform/biological sample/key outcomes
<i>Dietary intervention in normal weight individuals</i>				
Hooda <i>et al.</i> ⁽⁵⁶⁾	n 20 M; healthy adults	Randomised controlled crossover, no fibre, polydextrose (21 g/d); soluble maize fibre (21 g/d)	21 d	Pyrosequencing of V4 region of 16S rRNA gene/faeces. PCA showed gut microbial community shifts with fibre interventions
Ross <i>et al.</i> ⁽⁵⁷⁾	n 17; 11 F, 6 M	Randomised controlled crossover, WG(150/d) v. refined grain	2-week	Bacterial enumeration using FISH; <i>C. leptum</i> group increased in WG diet along with stool frequency
Costabile <i>et al.</i> ⁽⁵⁸⁾	n 31; 16 F, 15 M	Randomised controlled crossover, 48 g/d breakfast cereal of either 100% WG (11.8 g DF/100g; chosen after a pre-screening for bifidogenicity) or WB (27 g DF/100 g) <i>ad-lib</i> diet; energy composition was not the same; WG contained higher content of non-sugar carbohydrate	two 3-week periods	Bacterial group enumeration using FISH showed changes in several groups with both WG and WB diets and differences between diets. No change in faecal SCFA. Increase in fasting plasma ferulic acid with WB
Finley <i>et al.</i> ⁽⁵⁹⁾	n 40 pre-metabolic syndrome; n 40 controls; 20 F, 20 M	1/2 cup (130 g) of pinto bean puree v. chicken–noodle soup	4-week equilibration and 12-week intervention	Faecal bacterial species enumeration using FISH. Breath methane measured. No effect of bean consumption, except <i>E. limosum</i> levels decreased by 50%
Smith <i>et al.</i> ⁽⁶⁰⁾	n 18 M	Single-blind randomised crossover, self-managed with addition of seven experimental foods (bread, muffin, brownie, choc milk drink, muesli, pasta, mashed potatoes) with or without LKF; LKF diet provided 17–30 g additional fibre/d; mean fibre intake was 23 g/d on control and 45 g/d on LKF; LKF diet was sig. lower in starch	28-d ;3-d pooled faecal collection at end of each period	FISH of 16S rRNA genes/faeces with probes for total bacteria and specific groups. No difference in total bacteria, but changes in certain groups in response to treatment
Johnson <i>et al.</i> ⁽⁶¹⁾	n 38 M	Single-blind crossover, same as Smith <i>et al.</i> ⁽⁶⁰⁾ aforementioned	28-d; 3-d pooled faecal collection at end of each period	SCFA and bacterial enzyme activity in faeces. LKF altered bowel function parameters and decreased faecal pH. Faecal SCFA increased and β-glucuronidase activity decreased. No difference in faecal ammonia concentration
Tuohy <i>et al.</i> ⁽²⁵⁾	n 31; 17F, 14M	20 g FOS + 10 g PHGG/d v. placebo	crossover, 3-week feeding periods; two consecutive d stool sample mix	FISH of 16S rRNA genes. Increase in <i>Bifidobacterium</i> spp., but no difference in total bacteria between diets or other spp. enumerated. No change in faecal pH
Hylla <i>et al.</i> ⁽⁶²⁾	n 12; 5 F, 7 M	Randomised controlled crossover, high v. low resistant starch: amylo maize starch in bread, pasta, cake and biscuits (Hylon VII, Natl Starch)	4-week	Changes in breath hydrogen and several faecal parameters associated with bacterial activity (e.g. pH, certain SCFA, β-glucosidase, secondary bile acids)
<i>Dietary intervention in overweight/obese individuals</i>				
Weickert <i>et al.</i> ⁽⁶³⁾	n 69; 43 F, 26 M	Randomised controlled crossover, high cereal fibre 43 g/d, moderately high cereal/fibre/protein diets, 23% energy as protein, cereal fibre 26 g/d	18 weeks	FISH/Flow cytometry of faecal bacteria. No effect of diet on bacterial groups
Russell <i>et al.</i> ⁽²³⁾	n17 M; obese	Randomised controlled crossover. Maintenance diet (85 g protein, 116 g fat and 360 g carbohydrate/d) HPMC (139 g protein, 82 g fat and 181 g carbohydrate/d) HPLC (137 g protein, 143 g fat and 22 g carbohydrate/d)	7 d maintenance followed by 14-d intervention	FISH of 16S rRNA genes/GC-MS analysis of faecal water content/faeces. HPMC and HPLC diets resulted in increased proportions of branched-chain fatty acids and concentrations of phenylacetic acid and N-nitroso compounds. HPLC diet decreased proportion of butyrate in faecal SCFA concentrations, concomitant with reduction in Roseburia/Eubacterium rectale bacteria, and reduced fibre-derived, antioxidant phenolic acids

Table 1 (Continued)

Dietary intervention, reference and year	Sample size/population	Dietary intervention/dose of food agent	Treatment duration	Platform/biological sample/key outcomes
Duncan <i>et al.</i> ⁽⁶⁴⁾	n 19 M	Randomised controlled crossover, high (399 g carbs, 52%), medium (HPMC; 164 g/d, 35%) or low carb (HPLC; 24 g/d, 4%) <i>ad-lib</i> diet for 4 weeks; nonstarch polysaccharide levels in maintenance: 28 g, HPMC 12 g, HPLC 6 g	3-d maintenance/ 4-week experimental diet	Bacterial enumeration using FISH showed shifts in certain groups of bacteria. Faecal SCFA and ammonia lower during experimental periods than maintenance and proportions of some SCFA differed by diet
Ley <i>et al.</i> ⁽⁶⁵⁾	n 12; obese adults	Reduced fat about 30% of calories from fat or 25% of calories from carbohydrate; fibre was 10–15 g/d	1 year	Pyrosequencing of 16S <i>rRNA</i> gene. Ratio of Bacteroidetes to Firmicutes decreased over time associated with weight loss. No difference between diets

DF, dietary fibre; F, female; FISH, fluorescent *in-situ* hybridisation of 16S *rRNA* genes; FOS, fructo-oligosaccharide; HPMC, high-protein and moderate-carbohydrate; HPLC, diet and a high-protein and low-carbohydrate; LKF, lupin kernel fibre; M, male; PCA, principal components analysis; PHGG, partially hydrolysed guar gum; WB, wheat bran; WG, whole grain.

other endpoints beyond differences in metabolite profiles. Brauer *et al.*⁽³²⁾, interrogated the proteome in response to 2 weeks of a diet high in cruciferous vegetables, and assessed whether response differed by glutathione *S*-transferase (*GST*)*M1* genotype. GST enzymes metabolise a variety of exogenous compounds, including isothiocyanates from cruciferous vegetables, and the *GSTM1* variants resulting in a complete lack of gene product are common⁽²²⁾. Twenty-four distinct peaks were associated with cruciferous vegetable consumption compared with a fruit- and vegetable-free diet, two of which were identified that changed in a *GSTM1*-genotype-dependent manner. Another study provides an example of a novel use of omics to link metabolic phenotypes with dietary preferences. Taking a targeted approach, Rezzi *et al.*⁽³³⁾ used lipidomics to determine metabolites associated with chocolate ‘desiring’ or chocolate ‘indifferent’ preferences among individuals consuming 50 g/d chocolate or bread as a placebo. Heinzmann *et al.*⁽³⁴⁾ used metabolomics to study the stability of phenotypic response to diet through sequential dietary challenges. They found that inter-individual differences were often greater than differences within an individual in response to dietary modulation, providing evidence that individuals each have a unique metabolic phenotype. Moreover, intra-individual differences between consecutive dietary challenges were linked to differences in excretion of microbial co-metabolites suggesting flexibility in gut microbiome function in response to dietary modulation. As the authors point out, these differences illustrate the importance of assessing response to diet in the context of a crossover rather than parallel study design in order to move towards personalised nutrition. As a whole, these controlled feeding studies illustrate the potential for omics technology in characterising individual nutritional phenotypes, but make evident the challenges (i.e. compound identification, pathway analysis) that still exist.

How can we most effectively integrate omics data so as to be able to apply them towards personalised nutrition?

Given that cellular functions are carried out via orchestrated activities of multiplex components of biological systems, data from different omics platforms can shed light on cellular activities at different levels. Methods that integrate omics data from different molecular profiling studies, e.g. data from transcriptomics, proteomics or metabolomics studies, have the potential to provide new insight into how different components of biological systems interact with each other and form the basis of an individual’s health. Here, we provide an overview of available methods of data integration from multiple omics platforms, provide examples of each of different approaches, and discuss their advantages and limitations.

Current methods for integrative analysis of omics data from multiple data platforms can be broadly grouped into three categories. The first class of models, which we refer to as concordance analysis methods, studies concordance/correlation between two omics datasets, e.g. comparing the gene expression levels and proteomics datasets on the same

Table 2. Summary of human dietary intervention studies using metabolomic and proteomic platforms

Dietary intervention, reference and year	Sample size/population	Dietary intervention/dose of food agent	Treatment duration	Platform/biological sample/key outcomes
Fruits and vegetables				
May D, Navarro SL, Ruczinski I <i>et al.</i> (unpublished results) Metabolomic profiling of urine in response to a randomised, controlled feeding study of select fruits and vegetables and application to an observational study	<i>n</i> 10; 5 F, 5 M; healthy adults	Randomised controlled crossover, mixture of cruciferous vegetables, citrus fruits and soya (F&V) compared to fruit and vegetable-free diet (basal); 5 g/kg BW	2 weeks	Metabolomics/8 h fasting urine; more abundant in the F&V: markers of dietary intervention (e.g. crucifers, citrus and soya metabolites), fatty acids and niacin; more abundant in basal: riboflavin, several acylcarnitines and amino acid metabolites; differences in energy utilisation between diet treatments
Van Dorsten <i>et al.</i> ⁽⁶⁶⁾	<i>n</i> 58 (29 in each treatment arm); 25 F, 33 M hypertensive adults	Randomised double-blind placebo-controlled double-crossover, capsules containing a polyphenol rich mix of either red wine and red grape juice extracts (800 mg) or only red grape extract (800 mg), and placebo	4 weeks	Metabolomics/urine; 18 phenolic acids elevated after either polyphenol treatment including syringic acid, 3- and 4-hydroxyhippuric acid and 4-hydroxyhippuric acid and 4-hydroxymandelic acid
Brauer <i>et al.</i> ⁽³²⁾	<i>n</i> 36; 17 F, 19 M; <i>n</i> 42; 17 F, 25 M; healthy adults recruited based on <i>GSTM1</i> genotype (present or null)	2 separate randomised controlled crossover trials of mixed vegetables: (1) 436 g cruciferous; 90 g allium; and 270 g apiaceous; (2) 7 g/kg BW cruciferous; 14 g/kg BW cruciferous; 7 g/kg BW cruciferous + 4 g/kg BW apiaceous; both compared to fruit and vegetable-free diet (basal)	6 d; 2 weeks	Proteomics/8 h fasting serum; 24 distinct peaks associated with cruciferous vegetables; 20 associated with <i>GSTM1</i> genotype; joint study analysis showed 6 peaks changed in genotype-dependent manner; two identified as TTR and ZAG
Heinzmann <i>et al.</i> ⁽⁶⁷⁾	<i>n</i> 8; 7 F, 1 M healthy adults	Standardised mixed fruit meal (apple, orange, grapes and grapefruit; no dosage provided)	3 d	Metabolomics/urine; excretion of proline betaine, tartaric acid, hippuric acid and benzoic acid was increased compared to baseline
Walsh <i>et al.</i> ⁽⁶⁸⁾	<i>n</i> 21; 12 F, 9 M healthy adults	Non-controlled crossover, 2 d habitual diet; 2 d low phytochemical diet and 2 d high phytochemical diet (100 ml × 4 apple, carrot and strawberry drinks)	2 d	Metabolomics/fasting urine; higher excretion of hippurate and lower excretion of creatinine and methylhistidine discriminated the high phytochemical and habitual diets from the low phytochemical diet
Other plant foods				
Tulipani <i>et al.</i> ⁽⁶⁹⁾	<i>n</i> 42; adults with metabolic syndrome	Randomised parallel intervention, mixed nuts, 30 g/d and control	12 weeks	Metabolomics/urine; 20 potential markers of nut intake including fatty acid, phase II, microbially-derived phenolic, and serotonin metabolites
Llorach <i>et al.</i> ⁽⁷⁰⁾	<i>n</i> 24; healthy adults	Randomised blind placebo-controlled, encapsulated almond skin extract, 3.5 g	Single dose of ten capsules	Metabolomics/urine; 34 metabolites of almond skin including flavonoids, hydroxyphenylvalerolactone, 4-hydroxy-5-(phenyl)-valeric acid, hydroxyphenylpropionic acid, hydroxyphenylacetic acid and other phenolic acid conjugates
van Dorsten <i>et al.</i> ⁽⁷¹⁾	<i>n</i> 17 M; healthy adults	Randomised crossover, black tea (6 g/d), green tea (6 g/d) or caffeine (control)	2 d	Metabolomics/urine; green and black tea increased urinary excretion of hippuric acid and 1,3-dihydroxyphenyl-2- <i>O</i> -sulfate; greater increase in several citric acid cycle intermediates with green tea
Solanky <i>et al.</i> ⁽³¹⁾	<i>n</i> 9 F; pre-menopausal	Controlled, miso (50 g/d; <i>n</i> 6) or soya protein (60 g/d; <i>n</i> 3)	4 weeks	Metabolomics/urine; increased TMAO, methylamine, dimethyl amine, choline, creatine, glutamine (soya only) and glutamate (soya only), and decreased creatinine, hippurate, benzoate, citrate (miso only) and lactate (miso only)



Table 2 (Continued)

Dietary intervention, reference and year	Sample size/population	Dietary intervention/dose of food agent	Treatment duration	Platform/biological sample/key outcomes
Miscellaneous				
Rasmussen <i>et al.</i> ⁽⁷²⁾	<i>n</i> 77; 44 F, 33 M; overweight, adults	Randomised, high (23–28% of energy; <i>n</i> 42) or low (10–15% of energy; <i>n</i> 35) protein diet	6 months	Metabolomics/urinary; creatine increased with a high protein diet; citric acid increased with the low protein diet
Moazzami <i>et al.</i> ⁽⁷³⁾	<i>n</i> 17 M; prostate cancer patients	Randomised controlled crossover, whole grain rye, rye bran and refined white wheat product, control (485 g/d)	6 weeks	Metabolomics/fasting plasma; metabolites increased after rye bran included 3-hydroxybutyric acid, acetone, betaine, N,N-dimethylglycine and dimethyl sulfone
Heinzmann <i>et al.</i> ⁽³⁴⁾	<i>n</i> 7; 6 F, 1 M healthy adults	Controlled, various dietary challenges including mixed fruit (apple, orange, grapes and grapefruit), fish, wine and grapes, beef and fish	7 d	Metabolomics/urine; inter-individual metabolic differences were greater than effects of any single dietary challenge; differences to dietary challenges differed between individuals
Zivkovic <i>et al.</i> ⁽⁷⁴⁾	<i>n</i> 3; 1 F, 2 M; healthy adults	Standardised test beverages, 40% kcal needs; 230 g lactose-free milk, 227 g low-fat yogurt, 30 g 100% whey protein powder, 118 g banana and 22 g flax seed oil	Three single test beverages	Targeted lipidomics/plasma; serum fatty acid differences were greater among individuals than within; three metabolites discriminated individuals in ApoB fraction: TAG16 : <i>n</i> 7, TAG18 : 2 <i>n</i> 6 and phosphatidylcholine18 : 3 <i>n</i> 3
Llorach <i>et al.</i> ⁽⁷⁵⁾	<i>n</i> 10; 5 F, 5 M; Healthy adults	Randomised crossover, 40 g cocoa powder with water or 250 ml milk or 250 ml milk alone	Single test beverages	Metabolomics/urine; 27 cocoa-phytochemical metabolites identified after both cocoa-containing beverages
Bertram <i>et al.</i> ⁽⁷⁶⁾	<i>n</i> 28; 8-year-old boys	Randomised, 53 g/d protein from low-fat milk or low-fat meat	7 d	Metabolomics/urine and serum; urinary hippurate excretion was decreased with milk; urinary creatinine, histidine and urea was increased with meat
Rezzi <i>et al.</i> ⁽³³⁾	<i>n</i> 75 M; healthy adults	Controlled crossover, chocolate (50 g) or bread (placebo)	Single feedings	Lipidomics/urine and plasma; metabolic phenotypes associated with chocolate desiring or chocolate indifferent preferences
Stella <i>et al.</i> ⁽⁷⁷⁾	<i>n</i> 12 M; healthy adults	Randomised crossover, vegetarian (420 g/d), low-meat (60 g/d) or high-meat (420 g/d)	15 d	Metabolomics/urine; urinary excretion of carnitine, creatinine, taurine, TMAO, methyhistidine was increased with high-meat; <i>p</i> -hydroxyphenylacetate increased after vegetarian diet

BW, body weight; F, female; GSTM1, glutathione S-transferase M1; M, male; SMCSO, S-methyl-L-cysteine sulfoxide; TMAO, trimethylamine-N-oxide; TTR, transthyretin; ZAG, zinc α -2-glycoprotein.

set of subjects. The objective of such an approach is to identify genes/proteins/metabolites with an orchestrated activity in a given biological setting. To this end, methods of multivariate analysis, including different variations of principal component analysis, partial least squares, self-organising maps, as well as methods of network visualisation and analysis, have been used to assess the associations among multiple datasets. For instance, Hirai *et al.*⁽³⁵⁾ applied principal component analysis as well as self-organising maps to discover relationships between transcriptome and metabolome in *Arabidopsis*. In another study, Hirai *et al.*⁽³⁶⁾ analysed the network of gene-to-gene and gene-to-metabolite associations. More recently, Cao *et al.*⁽³⁷⁾ proposed a sparse partial least squares procedure for comparative analysis of data from two omics platforms and applied their method to data from cDNA and Affymetrix chips in NCI60 cancer cell lines.

Concordance analysis methods provide interesting information about components of biological systems that interact with each other in a given setting. Moreover, such analyses can lend themselves to better classificatory models based on a combination of biomarkers from different platforms. However, these approaches often provide limited new insight into the underlying biological mechanisms as omics data from different platforms often show low levels of correlation due to complex mappings of genes to proteins and metabolites, and various post-transcriptional events⁽³⁸⁾. Further, the underlying assumption in the majority of these methods is that omics measurements are obtained on the same set of individuals, or more formally, share a common dimension. Van Deun *et al.*⁽³⁹⁾ reviewed these different approaches for analysis of multiple omics data, in the setting where the datasets share a common set of features.

The second class of integrative models, which we refer to as sequential integration, includes methods that incorporate multiple sets of omics data in order to discover new biomarkers or delineate biological mechanisms of complex phenotypes. It uses multiple omics datasets, in a sequential manner, to further narrow down, or expand, the set of biomarkers. Sequential integration methods can exploit different methods of data analysis, from simple differential expression analysis to gene-set enrichment analysis or analysis of networks. In examples of such an approach, Putluri *et al.*⁽⁴⁰⁾ first identified the set of differentially active metabolites, and then used meta genomic data to identify pathways associated with prostate cancer progression. In another study, Putluri *et al.*⁽⁴¹⁾ coupled this approach with a concordance analysis based on metabolomics flux measurements to delineate pathways and biomarkers associated with bladder cancer. More recently, Imielinski *et al.*⁽⁴²⁾ used gene-set enrichment analysis coupled with the knowledge of biological networks and compared two sequential approaches, called 'gene-centric' and 'protein-centric,' in a study of molecular bases of breast cancer. In each of these approaches, the authors first evaluated the enrichment of biological pathways based on one source of data (transcriptomic or proteomic) and then filtered the set of identified pathways based on the second source of data. The authors also compared the results of these methods with a concordance-based approach, where

the pathways were identified based on gene and protein pairs that demonstrated orchestrated levels of activity.

Sequential integration methods offer an opportunity to gain new insight based on multiple sources of omics data. Moreover, these methods do not require the omics measurements to be necessarily observed for the same set of individuals. Finally, unlike methods of concordance analysis, which cannot be directly extended to analysis of more than two sets of omics data, sequential integration methods offer the flexibility of analysing multiple omics datasets. However, the power of these methods is clearly limited by the ability of the omics data chosen for the first stage of analysis to capture important biological mechanisms. As the study by Imielinski *et al.*⁽⁴²⁾ indicates, the results of the analysis can vary depending on the omics platform used for the first stage of analysis. This sensitivity to the order of analysis can potentially hinder the applications of sequential integration methods, and additional studies are needed to determine whether data-driven criteria can be developed to assess the optimal order of analysis in these methods.

The final group of omics integration techniques, which we refer to as concurrent integration methods, includes emerging approaches that attempt to address some of the shortcomings of the afore-mentioned two sets of approaches. Similar to sequential integration methods, concurrent integration methods try to exploit the information content of multiple sets of omics data. However, these methods often include measures of activity of biological pathways, or their components, based on multiple omics data. This is often achieved by defining a combined score for the activity of each pathway based on activities of its members measured by different omics datasets. Poisson *et al.*⁽⁴³⁾ compared the performance a number of methods for combining data from multiple omics platforms, by considering different summary measures defined based on individual test statistics, with methods based on a single omics data source and show that the integrative approaches can improve the power of the analysis. In a recent study, Jauhainen *et al.*⁽⁴⁴⁾ proposed a multivariate approach, using a mixed linear model, to assess the association of transcriptomics and metabolomics measurements with cancer progression. The proposed model requires measurements to be observed on the same set of samples, but offers the potential for discovering novel biological mechanisms, as well as biomarker identification. On the other hand, Shojaie *et al.* (A Shojaie, K Panzitt, N Putluri *et al.*, unpublished results) propose a network-based method, based on the NetGSA method⁽⁴⁵⁾, for integrating multiple sources of omics data, which can be applied to data from different samples. This procedure does not lend itself directly to selection of biomarkers, and follow-up analyses are needed to determine which components of the selected pathways should be used as biomarkers.

Concurrent integrative methods have also been proposed for gaining insight into biological mechanisms in the cell. An example of such an approach includes the proposal of Shojaie *et al.* (A Shojaie, A Jauhainen, M Kallitsis and G Michailidis, unpublished results) to integrate perturbation screens and steady-state gene expression profiles for discovering causal genetic regulatory mechanisms. In this

Table 3. Summary of methods for integrative analysis of multiple omics datasets

Integration approach	Reference	Methodology/tools	Omics data
(1) Concordance analysis	Hirai <i>et al.</i> ⁽³⁵⁾	PCA, SOM	Transcriptome and metabolome in <i>Arabidopsis</i>
	Hirai <i>et al.</i> ⁽³⁶⁾	Network analysis	Transcriptome and metabolome in <i>Arabidopsis</i>
	Le Cao <i>et al.</i> ⁽³⁷⁾	Sparse PLS	cDNA and mRNA in NCI60 cancer cell lines
	Van Deun <i>et al.</i> ⁽³⁹⁾	Multiple methods	Comparative analysis of integration methods assuming data on the same subjects
(2) Sequential integration	Putluri <i>et al.</i> ⁽⁴⁰⁾	DE, OCM	Metabolomics, meta-genomics in prostate cancer
	Putluri <i>et al.</i> ⁽⁴¹⁾	DE, OCM, CA, PLS	Metabolomics abundance and flux data, meta-genomics in bladder cancer
	Imielinski <i>et al.</i> ⁽⁴²⁾	GSEA, network analysis	Transcriptomics, proteomics in breast cancer
(3) Concurrent integration	Poisson <i>et al.</i> ⁽⁴³⁾	DE, <i>P</i> -value weighting, GSEA	Transcriptomics, metabolomics
	Jauhainen <i>et al.</i> ⁽⁴⁴⁾	Sparse mixed linear model	Transcriptomics and metabolomics in cancer
	Shojaie A, Panzitt K, Putluri N <i>et al.</i> (unpublished results)	NetGSA, GSEA, rank-based integration	Transcriptomics and metabolomics in prostate cancer
	A network-based Integrative approach to study the role of metabolic pathways in prostate cancer progression		

DE, differential analysis; GSEA, gene-set enrichment analysis; CA, correlation analysis; PCA, principal component analysis; SOM, self-organising maps; PLS, partial least squares; OCM, Oncomine concept mapping; NetGSA, network-based gene-set analysis.

study, the authors compare their proposed integrative approach with state-of-the-art methods based on a single source of omics data, and show superior estimates of regulatory networks can be obtained that by combining multiple omics data. Table 3 summarises the different classes of integration methods.

Novel biomedical technologies continue to improve the quality of the omics data, as well as to reduce the cost of obtaining such data. In nutrition studies, biological experiments now generate multiple sources of omics data including transcriptomic, proteomic, metabolomic and gut microbial community measurements. The main challenge is now integrating such measurements in a systematic way, in order to provide a holistic view of biological systems. As more and more measurements become available, the complexity of the analysis, i.e. the number of variables in statistical models, increases. This poses additional challenges for design of trials, and necessitates the use of advanced statistical models appropriate for analysis of high-dimensional problems. A potential solution for this challenge is to incorporate available biological knowledge, including information on biological pathways and genetic, protein interaction and metabolic networks. Incorporating biological information can both reduce the dimensionality of the problem, and also improve the power and reproducibility of analysis methods.

A way forward to personalised nutrition

There is still a lot of effort needed to establish a robust health phenotype framework on which to develop personalised dietary recommendations. The improving omic technologies and the ability to integrate various

omics platforms in a systematic fashion will facilitate providing a holistic view of cellular functions related to healthy phenotypes; however, the characterisation of the contribution of diet to the biochemical and metabolic parameters associated with healthy phenotypes would benefit from systematic evaluation under controlled conditions in well-described groups of individuals. Controlled human feeding studies are a useful experimental setting in which to conduct this work. Nonetheless, these types of studies are expensive and funding multiple, new large-scale dietary interventions that capture a variety of dietary patterns and intakes is likely to be prohibitive.

An efficient and effective way to develop some of the necessary omics databases under experimental conditions may be to take a collaborative approach, leveraging existing samples from previously conducted human interventions. Stored samples from controlled feeding studies are stashed away in freezers around the globe and in many cases are well characterised and ideal for further omic analysis. Statistical techniques for integrating multiple omics data from a common platform but different study populations, i.e. meta-analysis techniques, already exist; they improve statistical power by integrating samples from multiple related studies^(46–54) and also allow for testing of reproducibility of results across studies. Looking towards future studies, the adoption of standardised sample and metadata collection protocols would allow for easier pooling of data across studies.

Overall, the careful collection and integration of omics data from controlled dietary interventions may provide us with the data necessary to successfully move towards a goal of more personalised dietary recommendations. Nonetheless, even with the generation of expansive, integrated datasets that allow for in-depth characterisation

of health phenotypes, several factors need to be considered if personalised nutrition is to move towards being a part of routine health practice. Adherence to dietary recommendations for chronic disease prevention at the population level, such as those of national and international associations (e.g. US Department of Agriculture, World Cancer Research Fund, American Heart Association) is associated with lower risk of chronic disease; for example, greater adherence to the 2005 US Dietary Guidelines was inversely associated with risk of CHD, stroke, diabetes and total cancer.⁽⁵⁵⁾ In theory, tailored recommendations may be an improvement over general, population-based dietary recommendations; however, whether more extensive phenotyping, beyond current approaches, is cost-effective in promoting health and preventing disease will need to be determined. Further, in practice, finding individualised approaches that facilitate and maintain desired dietary behaviour on the heels of a personalised diet prescription for health will likely remain an ongoing challenge for nutrition practitioners.

Acknowledgements

This work was supported in part by Fred Hutchinson Cancer Research Center, NCI U01 CA162077 and US NSF DMS-1161565. All authors declare no conflict of interest. The paper was written by J. W. L., S. L. N., M. A. J. H., and A. S. All authors read the draft critically and J. W. L. had the responsibility for final content.

References

1. Jones DP, Park Y & Ziegler TR (2012) Nutritional metabolomics: progress in addressing complexity in diet and health. *Annu Rev Nutr* **32**, 183–202.
2. Zeisel SH, Freake HC, Bauman DE *et al.* (2005) The nutritional phenotype in the age of metabolomics. *J Nutr* **135**, 1613–1616.
3. Lampe JW & Potter JD (2006) Genetic variation, diet and disease susceptibility. In *Gene-Environment Interactions: Fundamentals of Ecogenetics*, pp. 321–350 [LG Costa and DL Eaton, editors]. Hoboken, NJ: John Wiley & Sons, Inc.
4. Bryan AD & Hutchison KE (2012) The role of genomics in health behavior change: challenges and opportunities. *Public Health Genomics* **15**, 139–145.
5. Gieger C, Geistlinger L, Altmaier E *et al.* (2008) Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet* **4**, e1000282.
6. Suhre K, Wallaschofski H, Raffler J *et al.* (2011) A genome-wide association study of metabolic traits in human urine. *Nat Genet* **43**, 565–569.
7. Guarner F & Malagelada JR (2003) Gut flora in health and disease. *Lancet* **361**, 512–519.
8. Qin J, Li R, Raes J *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65.
9. Gill SR, Pop M, Deboy RT *et al.* (2006) Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355–1359.
10. Arumugam M, Raes J, Pelletier E *et al.* (2011) Enterotypes of the human gut microbiome. *Nature* **473**, 174–180.
11. Donohoe DR & Bultman SJ (2012) Metaboloeigenetics: interrelationships between energy metabolism and epigenetic control of gene expression. *J Cell Physiol* **227**, 3169–3177.
12. Donohoe DR, Garge N, Zhang XX *et al.* (2011) The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab* **13**, 517–526.
13. Ordovas JM & Mooser V (2006) Metagenomics: the role of the microbiome in cardiovascular diseases. *Curr Opin Lipidol* **17**, 157–161.
14. Qin JJ, Li YR, Cai ZM *et al.* (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60.
15. Marchesi JR, Dutilh BE, Hall N *et al.* (2011) Towards the human colorectal cancer microbiome. *PLoS One* **6**, e20447.
16. Plottel CS & Blaser MJ (2011) Microbiome and malignancy. *Cell Host Microbe* **10**, 324–335.
17. Willing BP, Dicksved J, Halfvarson J *et al.* (2010) A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139**, 1844–1854.
18. Lampe JW (2004) Nutrition and cancer prevention: small-scale human studies for the 21st century. *Cancer Epidemiol Biomarkers Prev* **13**, 1987–1988.
19. Navarro SL, Chang J, Peterson S *et al.* (2009) Modulation of human serum glutathione S-transferase-A1/2 concentration by cruciferous vegetables in a controlled feeding study is influenced by GSTM1 and GSTT1 genotypes. *Cancer Epidemiol Biomarkers Prev* **18**, 2974–2978.
20. Cross AJ, Major JM, Sinha R (2011) Urinary biomarkers of meat consumption. *Cancer Epidemiol Biomarkers Prev* **20**, 1107–1111.
21. Neuhouser ML, Schwarz Y, Wang C *et al.* (2012) A low-glycemic load diet reduces serum C-reactive protein and modestly increases adiponectin in overweight and obese adults. *J Nutr* **142**, 369–374.
22. Navarro SL, Chen Y, Li L *et al.* (2011) UGT1A6 and UGT2B15 polymorphisms and acetaminophen conjugation in response to a randomized, controlled diet of select fruits and vegetables. *Drug Metab Dispos* **39**, 1650–1657.
23. Russell WR, Gratz SW, Duncan SH *et al.* (2011) High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health. *Am J Clin Nutr* **93**, 1062–1072.
24. Li F, Hullar MA, Schwarz Y *et al.* (2009) Human gut bacterial communities are altered by addition of cruciferous vegetables to a controlled fruit- and vegetable-free diet. *J Nutr* **139**, 1685–1691.
25. Tuohy KM, Kolida S, Lustenberger AM *et al.* (2001) The prebiotic effects of biscuits containing partially hydrolysed guar gum and fructo-oligosaccharides—a human volunteer study. *Br J Nutr* **86**, 341–348.
26. Faust K & Raes J (2012) Microbial interactions: from networks to models. *Nat Rev Microbiol* **10**, 538–550.
27. Faust K, Sathirapongsasuti JF, Izard J *et al.* (2012) Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol* **8**, e1002606.
28. Lozupone C, Faust K, Raes J *et al.* (2012) Identifying genomic and metabolic features that can underline early successional and opportunistic lifestyles of human gut symbionts. *Genome Res* **22**, 1974–1984.
29. Bolca S, Van de Wiele T & Possemiers S (2012) Gut metabolites govern health effects of dietary polyphenols. *Curr Opin Biotechnol* (Epublication ahead of print version).
30. Niculescu MD, Pop EA, Fischer LM *et al.* (2007) Dietary isoflavones differentially induce gene expression changes in lymphocytes from postmenopausal women who form equal

- as compared with those who do not. *J Nutr Biochem* **18**, 380–390.
31. Solanky KS, Bailey NJ, Beckwith-Hall BM *et al.* (2005) Biofluid 1H NMR-based metabonomic techniques in nutrition research - metabolic effects of dietary isoflavones in humans. *J Nutr Biochem* **16**, 236–244.
 32. Brauer HA, Libby TE, Mitchell BL *et al.* (2011) Cruciferous vegetable supplementation in a controlled diet study alters the serum peptidome in a GSTM1-genotype dependent manner. *Nutr J* **10**, 11.
 33. Rezzi S, Ramadan Z, Martin FP *et al.* (2007) Human metabolic phenotypes link directly to specific dietary preferences in healthy individuals. *J Proteome Res* **6**, 4469–4477.
 34. Heinzmann SS, Merrifield CA, Rezzi S *et al.* (2012) Stability and robustness of human metabolic phenotypes in response to sequential food challenges. *J Proteome Res* **11**, 643–655.
 35. Hirai MY, Yano M, Goodenowe DB *et al.* (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **101**, 10205–10210.
 36. Hirai MY, Klein M, Fujikawa Y *et al.* (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. *J Biol Chem* **280**, 25590–25595.
 37. Lê Cao KA, Martin PG, Robert-Granié C *et al.* (2009) Sparse canonical methods for biological data integration: application to a cross-platform study. *BMC Bioinformatics* **10**, 34.
 38. Gygi SP, Rochon Y, Franz BR *et al.* (1999) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**, 1720–1730.
 39. Van Deun K, Smilde AK, van der Werf MJ *et al.* (2009) A structured overview of simultaneous component based data integration. *BMC Bioinformatics* **10**, 246.
 40. Putluri N, Shojaie A, Vasu VT *et al.* (2011) Metabolomic profiling reveals a role for androgen in activating amino acid metabolism and methylation in prostate cancer cells. *PLoS One* **6**, e21417.
 41. Putluri N, Shojaie A, Vasu VT *et al.* (2011) Metabolomic profiling reveals potential markers and bioprocesses altered in bladder cancer progression. *Cancer Res* **71**, 7376–7386.
 42. Imielinski M, Cha S, Rejtar T *et al.* (2012) Integrated proteomic, transcriptomic, and biological network analysis of breast carcinoma reveals molecular features of tumorigenesis and clinical relapse. *Mol Cell Proteomics* **11**, M111 014910.
 43. Poisson LM, Taylor JM & Ghosh D (2011) Integrative set enrichment testing for multiple omics platforms. *BMC Bioinformatics* **12**, 459.
 44. Jauhainen A, Nerman O, Michailidis G *et al.* (2012) Transcriptional and metabolic data integration and modeling for identification of active pathways. *Biostatistics* **13**, 748–761.
 45. Shojaie A & Michailidis G (2010) Network enrichment analysis in complex experiments. *Stat Appl Genet Mol Biol* **9**, Article22, 1–34.
 46. Coviello AD, Haring R, Wellons M *et al.* (2012) A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci implicated in sex steroid hormone regulation. *PLoS Genet* **8**, e1002805.
 47. Qayyum R, Snively BM, Ziv E *et al.* (2012) A meta-analysis and genome-wide association study of platelet count and mean platelet volume in african americans. *PLoS Genet* **8**, e1002491.
 48. Nyholt DR, Low SK, Anderson CA *et al.* (2012) Genome-wide association meta-analysis identifies new endometriosis risk loci. *Nat Genet* **44**, 1355–1359.
 49. Houlston RS, Webb E, Broderick P *et al.* (2008) Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* **40**, 1426–1435.
 50. Zeggini E, Scott LJ, Saxena R *et al.* (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* **40**, 638–645.
 51. Cooper JD, Smyth DJ, Smiles AM *et al.* (2008) Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat Genet* **40**, 1399–1401.
 52. Wirapati P, Sotiriou C, Kunkel S *et al.* (2008) Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* **10**, R65.
 53. Rhodes DR, Yu J, Shanker K *et al.* (2004) Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci USA* **101**, 9309–9314.
 54. Rhodes DR, Barrette TR, Rubin MA *et al.* (2002) Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* **62**, 4427–4433.
 55. Chiuvè SE, Fung TT, Rimm EB *et al.* (2012) Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr* **142**, 1009–1018.
 56. Hooda S, Boler BM, Serao MC *et al.* (2012) 454 pyrosequencing reveals a shift in fecal microbiota of healthy adult men consuming polydextrose or soluble corn fiber. *J Nutr* **142**, 1259–1265.
 57. Ross AB, Bruce SJ, Blondel-Lubrano A *et al.* (2011) A whole-grain cereal-rich diet increases plasma betaine, and tends to decrease total and LDL-cholesterol compared with a refined-grain diet in healthy subjects. *Br J Nutr* **105**, 1492–1502.
 58. Costabile A, Klinder A, Fava F *et al.* (2008) Whole-grain wheat breakfast cereal has a prebiotic effect on the human gut microbiota: a double-blind, placebo-controlled, crossover study. *Br J Nutr* **99**, 110–120.
 59. Finley JW, Burrell JB & Reeves PG (2007) Pinto bean consumption changes SCFA profiles in fecal fermentations, bacterial populations of the lower bowel, and lipid profiles in blood of humans. *J Nutr* **137**, 2391–2398.
 60. Smith SC, Choy R, Johnson SK *et al.* (2006) Lupin kernel fiber consumption modifies fecal microbiota in healthy men as determined by rRNA gene fluorescent *in situ* hybridization. *Eur J Nutr* **45**, 335–341.
 61. Johnson SK, Chua V, Hall RS *et al.* (2006) Lupin kernel fibre foods improve bowel function and beneficially modify some putative faecal risk factors for colon cancer in men. *Br J Nutr* **95**, 372–378.
 62. Hylla S, Gostner A, Dusel G *et al.* (1998) Effects of resistant starch on the colon in healthy volunteers: possible implications for cancer prevention. *Am J Clin Nutr* **67**, 136–142.
 63. Weickert MO, Arafat AM, Blaut M *et al.* (2011) Changes in dominant groups of the gut microbiota do not explain cereal-fiber induced improvement of whole-body insulin sensitivity. *Nutr Metab (Lond)* **8**, 90.
 64. Duncan SH, Belenguer A, Holtrop G *et al.* (2007) Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Appl Environ Microbiol* **73**, 1073–1078.
 65. Ley RE, Turnbaugh PJ, Klein S *et al.* (2006) Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023.
 66. van Dorsten FA, Peters S, Gross G *et al.* (2012) Gut microbial metabolism of polyphenols from black tea and red wine/grape juice is source-specific and colon-region dependent. *J Agric Food Chem* **60**, 11331–11342.



67. Heinzmann SS, Brown IJ, Chan Q *et al.* (2010) Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *Am J Clin Nutr* **92**, 436–443.
68. Walsh MC, Brennan L, Pujos-Guillot E *et al.* (2007) Influence of acute phytochemical intake on human urinary metabolomic profiles. *Am J Clin Nutr* **86**, 1687–1693.
69. Tulipani S, Llorach R, Jauregui O *et al.* (2011) Metabolomics unveils urinary changes in subjects with metabolic syndrome following 12-week nut consumption. *J Proteome Res* **10**, 5047–5058.
70. Llorach R, Garrido I, Monagas M *et al.* (2010) Metabolomics study of human urinary metabolome modifications after intake of almond (*Prunus dulcis* (Mill.) D.A. Webb) skin polyphenols. *J Proteome Res* **9**, 5859–5867.
71. Van Dorsten FA, Daykin CA, Mulder TP *et al.* (2006) Metabonomics approach to determine metabolic differences between green tea and black tea consumption. *J Agric Food Chem* **54**, 6929–6938.
72. Rasmussen LG, Winning H, Savorani F *et al.* (2012) Assessment of the effect of high or low protein diet on the human urine metabolome as measured by NMR. *Nutrients* **4**, 112–131.
73. Moazzami AA, Zhang JX, Kamal-Eldin A *et al.* (2011) Nuclear magnetic resonance-based metabolomics enable detection of the effects of a whole grain rye and rye bran diet on the metabolic profile of plasma in prostate cancer patients. *J Nutr* **141**, 2126–2132.
74. Zivkovic AM, Wiest MM, Nguyen U *et al.* (2009) Assessing individual metabolic responsiveness to a lipid challenge using a targeted metabolomic approach. *Metabolomics* **5**, 209–218.
75. Llorach R, Urpi-Sarda M, Jauregui O *et al.* (2009) An LC-MS-based metabolomics approach for exploring urinary metabolome modifications after cocoa consumption. *J Proteome Res* **8**, 5060–5068.
76. Bertram HC, Hoppe C, Petersen BO *et al.* (2007) An NMR-based metabonomic investigation on effects of milk and meat protein diets given to 8-year-old boys. *Br J Nutr* **97**, 758–763.
77. Stella C, Beckwith-Hall B, Cloarec O *et al.* (2006) Susceptibility of human metabolic phenotypes to dietary modulation. *J Proteome Res* **5**, 2780–2788.