Metabonomics and Metabolomics

David J. Grainger[1] *and Jeremy K. Nicholson*[2]
[1] *Department of Medicine, University of Cambridge, Cambridge, UK*
[2] *Department of Biological Chemistry, Imperial College, London, UK*

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.

ISBN: 3-527-30550-5

Keywords

Chromatography
A range of methods designed to separate the molecular components of a complex mixture on the basis of their physical properties.

Mass Spectrometry
An analytical chemistry technique that can be used to identify molecules by fragmenting them into ions, which are then separated on the basis of mass.

Nuclear Magnetic Resonance
The context-dependent resonance of various atomic nuclei (such as the hydrogen nucleus) when placed in a magnetic field, which can be used to obtain structural information about the molecular components of a solution.

Regression Modeling
Mathematical methods designed to predict one (or more) variables from the values of many other measured variables.

Metabonomics is the study of systemic biochemical profiles and regulation of function in whole organisms by analyzing biofluids and tissues. Like genomics (the study of the complete repertoire of genes in an organism) and proteomics (the study of the protein complement of a tissue or cell), metabonomics can provide a holistic overview of the current physiological status of an organism, and its response to external stressors. Here we review the technological approaches to generating metabolic profiles, highlighting the advantages and disadvantages of each methodology, as well as the various strategies for extracting useful conclusions from the very large datasets that can be generated by such profiling. Metabonomics can be applied to a wide range of biological applications, including predictive toxicology, probing the physiology of disease both in animal models and in man, and to make clinically useful diagnoses of disease. With examples of each of these applications, we illustrate the potential of metabonomics to contribute to our understanding of complex biological systems in a post-genomic era where we understand many of the components of living systems, but few of their dynamic interactions.

1 What is Metabonomics?

The suffix ''omics'' is now routinely applied in many fields to the holistic study of an entire system, as opposed to a reductionist description of each of its parts independently. Thus, metabonomics is the name given to the holistic study of metabolic systems in living organisms. In principle, a metabolic profile is therefore a simple list of all the low molecular weight metabolites (such as sugars, amino acids, and lipids) present in a biological system, together with the concentration of each metabolite present (the generation of such profiles is one of the definitions in use for metab*o*lomics). Clearly, such a profile is analogous to a genomic profile (a list of all the genes composing an organism, perhaps with their levels of expression also) or a proteomic profile (a list of the proteins in an organism).

Like genomics and proteomics, however, metabo*n*omics is also much more than a simple list. The metabolic profile of a particular biological sample is just a snapshot of a complex, dynamic network that reflects the physiological activity of the organism. Enzymes are rapidly interconverting metabolites; new compounds are being absorbed from the environment; waste products are being excreted. The science of metabonomics, therefore, is not only about capturing metabolic profiles (described in Sect. 2 below) but also about extracting an understanding of the underlying biological system from the resulting dataset (described in Sect. 3). In particular, metabonomics is a global metabolic regulation approach based on understanding complex system behavior, designed to reveal the response of an organism to an external stressor or stimulus.

The relationship between metabonomics and other systems biology disciplines is illustrated in Fig. 1. Genetic information (coupled with the pre-existing levels of various other proteins) is the major determinant of the mRNA (or transcriptomic) profile, which in turn is a key determinant of the proteomic profile. Proteins (in the form of enzymes) are an important determinant of the metabolic profile, which, in turn, feeds back to modulate gene expression patterns. This ''homeostatic loop'' is then modulated by external inputs from the environment. The major input of low molecular weight compounds from the diet can have a significant effect on the metabonomic profile, which subsequently affects the gene expression and protein profiles of the organism. Other environmental factors can also affect the metabolic profile (for example, the amount of light determines the rate of vitamin D formation), as well as directly altering gene expression (the amount of exercise modulates skeletal muscle gene expression and ultimately protein content). This position at the interface between genetic and environmental determinants makes metabolic profiling a uniquely powerful tool for probing the dynamic physiological status of an organism.

Compared with genomic and proteomic profiles, the metabonomic profile is also more dynamic, reflecting the current physiological status of the organism as well as its future behavior. Polymorphisms, in particular genes (a component of the genomic profile), may allow the risk of developing a disease to be estimated, but they cannot determine whether the organism is suffering from that disease at a given point in time.

To properly understand the metabolic network of a multicellular organism would require continuous measurement of the

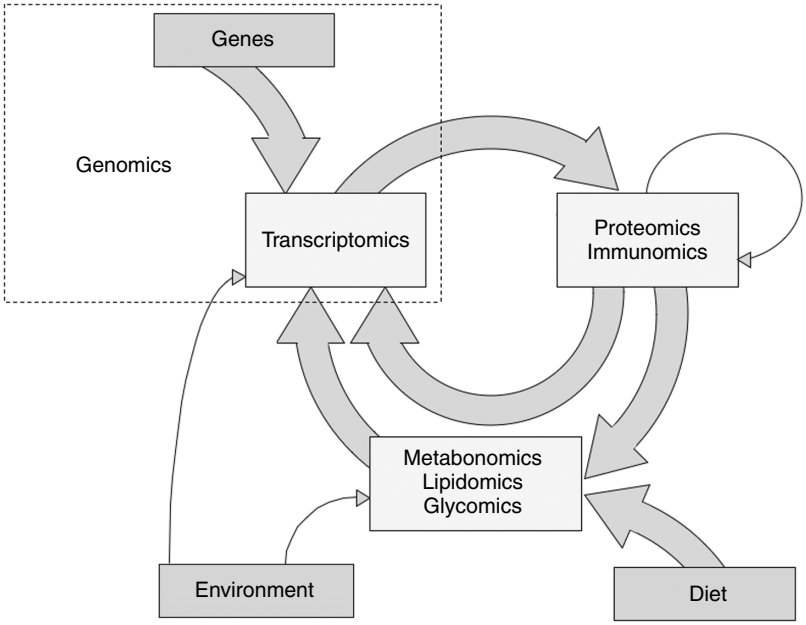

Fig. 1 Relationship between the ''omics.'' Much of the variation in protein levels (measured in proteomics) is due to variations in the expression of the mRNA encoding that protein (measured in transcriptomics, a subset of genomics). In turn, the variation in the levels of metabolites (measured in metabonomics) is determined by the levels of various enzymes (proteins). Metabolites can then feedback and regulate gene expression patterns, closing a ''homeostatic loop.'' The environment interacts with this homeostatic loop primarily through the influence of diet on the metabolic profile, although other environmental factors can also have a direct effect on both metabolism and gene expression (for example, exposure to UV light directly affects vitamin D3 levels, and exercise can affect gene expression patterns).

levels of all the metabolites present in all of the cells and tissues that compose the organism. Such an ''ideal'' metabolic profile is unlikely to be practicable attainable (at least in the near future). As a result, practical metabonomics involves selection of both a sample of the whole organism (for example, a blood specimen) and a time point (or series of time points) at which to make the observations. Clearly, the extent to which one can hope to understand the metabolism of the organism as a whole from a (potentially poorly representative) sample is unclear and care should be taken in drawing broad conclusions from limited measurements. Similarly, available analytical chemistry techniques (such as spectroscopy or chromatography) do not allow accurate measurements of the levels of every low molecular weight compound present in a given biological specimen. These practical limitations are illustrated in Fig. 2.

Even with the current practical limitations, metabonomics is a powerful new tool for studying complex biological systems. It has already been used successfully to monitor the physiological response to xenobiotics (such as new pharmaceuticals under development), and its use in

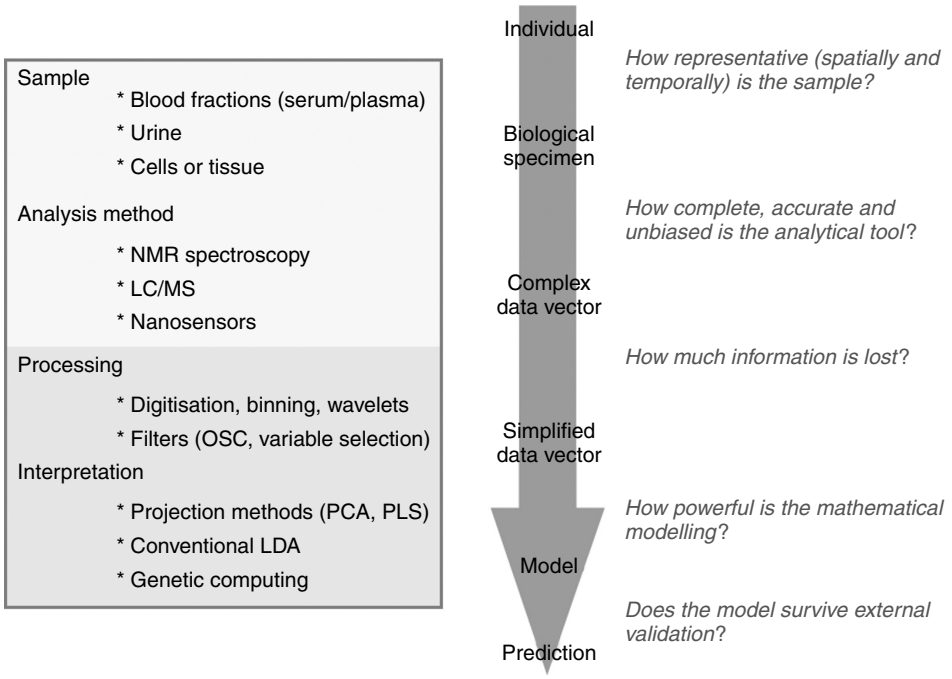

Fig. 2 A generic ''omics'' experiment. The aim of ''omics'' biology is to make measurements on individuals and then deduce predictive rules about the organism. In order to go from individual observations to a prediction, various steps must be followed: a sample must be taken on which many measurements are made to generate a complex data vector. Various mathematical modeling tools are then used to build a regression model which can be used to make a prediction that can be validated. Examples of each step in a typical metabonomics experiment are shown in the left panel. Various assumptions made at each step are listed in the right panel.

toxicology is increasing. Metabonomics can also be used to diagnose the presence of diseases, both for those where in-born errors of metabolism are responsible for the symptoms, and also in diseases where metabolic disregulation is less obviously involved in the pathogenesis of the disease. Using a metabolic profile to diagnose disease is not just useful in the clinic: it can also provide important new information about the physiological processes that are misregulated in the disease, which might ultimately assist in the search for new treatments.

The ''homeostatic loop'' illustrated in Fig. 1 also emphasizes the importance of integrating the genetic, protein, and metabolic profiles if we are to maximize our understanding of the organism. The boundaries between the ''omics'' represent technical limits in our methodologies for making measurements rather than being any useful dividing line between the applications of the information that has been generated. Fortunately, the profiles generated can be merged (at least in principle, although the bioinformatics challenges in doing so are significant) yielding a composite

(or ''multi-omics'') profile of the organism. Any question partially addressed by the separate genomic, proteomic, and metabonomic profiles will likely be more fully answered through the careful construction and analysis of such ''multi-omics'' profiles.

2 Methods for Generating a Metabonomic Profile

2.1 Criteria for Judging Metabonomic Profiling Methods

A range of different analytical chemistry approaches have been used to generate metabonomic profiles. Unfortunately, none of the profiles generated even approximate to the ''ideal'' profile, and it is necessary to make an informed choice of analytical tool depending on various trade offs. There are three important criteria that contribute to the ''ideal'' profile:

1. *Completeness*. If we assume that the ''ideal'' profile consists of a list of all the different low molecular weight compounds present in a biological sample, then any profiling method can be judged on the fraction of all the metabolites present that contribute to the profile. In absolute terms, this can be difficult to assess, since without a ''gold standard'' complete profile it is impossible to know what components have been missed. In practice, however, applying multiple analytical approaches to the same sample soon throws up examples of components missed by the other techniques.

Two factors contribute to the completeness of a profile. Firstly, the general sensitivity limit of the technique sets the threshold below which no components are detected. While an ''ideal'' profile would include even components present as just a single molecule, practicality suggests that components present at such low levels are unlikely to have a biologically significant effect, and that a profile with sensitivity threshold in the pM or nM range would be adequate for all but the most demanding applications. Unfortunately, several techniques are more insensitive still, and will miss components that are biologically relevant. It is worth noting, however, that sensitivity can be a double-edged sword – many of the low abundance metabolic components are derived from symbiotic or xenobiotic organisms (such as gut microflora) not directly related to mammalian metabolism. It is unclear whether gathering such enlarged metabolic datasets using highly sensitive analytical approaches is useful – indeed, it may simply make the task of extracting a meaningful picture from the resulting dataset more difficult (see Sect. 3).

The second factor is the invisibility of particular compounds, or more likely classes of compounds, to a particular analytical technique. An obvious example would be the inability of non-volatile components to contribute to a gas chromatograph. Less obvious might be the inability of a technique to separate or distinguish components with closely related structures, such that a single ''entry'' in the profile is in fact a composite measure of two or more compounds.

2. *Bias*. The ''ideal'' profile is not merely a list of the components present in the sample but also an indicator of the relative amounts of each component. Thus, if certain components are more readily detected than others (on a molar basis) the analytical technique will display a bias, suggesting some components are present at relatively higher levels than they actually are.
3. *Cost*. The resource implications of any data-gathering exercise must be properly considered as part of the scientific experimental design. This is particularly true of any nonselective, high data density ''omics'' experiment, where the amount of understanding about the system which is ultimately gained will likely depend on how many different (that is, uncorrelated) components of the system are analyzed, rather than whether any particular class of components (be it metabolites, proteins or genes) has been exhaustively investigated. Consequently, the resource implications of selecting any given analytical methodology rightly forms a part of the experimental design: does the additional information gained by adding a particular technique to the portfolio of analyses to be performed on a given sample add sufficient uncorrelated variables to justify the resources employed, or could the same resources generate more uncorrelated information through application of a different technique?

No technique currently available is complete, unbiased, and inexpensive, but each has certain advantages for particular applications. Application of multiple techniques to the same sample may improve completeness and reduce bias, but only at increased resource implication.

2.2 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR (nuclear magnetic resonance) spectroscopy has been widely used to generate metabonomic profiles, particularly of serum and urine samples. The primary advantages of NMR spectroscopy are the intrinsic reproducibility of the generated spectrum and the complete lack of bias. Across the information-dense region of the spectrum the coefficient of variation between replicate measures made on different days is below 1%. Reproducibility of this nature allows even small differences between profiles to be interpreted as significant, increasing the power of the experiment, particularly when relatively small numbers of profiles are being compared.

The NMR spectrum depends on the context-dependent resonance of hydrogen nuclei within the various molecules that compose the biological sample. As a result, any molecular structure containing at least one hydrogen nucleus is in principle represented within the spectrum. Since all biological molecules fall into this category, essentially every metabolite can contribute to an NMR-derived metabonomic profile. Furthermore, the intensity of the signal due to each hydrogen nucleus is of the same strength irrespective of its molecular context. Consequently, there is absolutely no bias in the estimated relative amounts of each of the metabolites detected.

Despite this lack of bias, NMR spectroscopy cannot be considered to generate a complete metabonomic profile because of the inherent insensitivity of the approach. Although the high reproducibility

allows even very small peaks to be distinguished from baseline noise, these peaks still represent relatively abundant molecular components of the sample. Although the absolute sensitivity cutoff varies depending on the particular implementation of the technique (for example, the use of cryoprobes significantly improves the sensitivity with nanogram quantities of compounds detected), nevertheless, very low abundance molecules are difficult to detect. Since many biological molecules of importance (such as vitamins and signaling molecules like prostaglandins or cyclicAMP) rarely achieve concentrations above the nM range, they are effectively absent from NMR-derived metabonomic datasets.

Although NMR spectroscopy is the tool of choice for most chemical structure determination problems, it also struggles to distinguish certain classes of closely related molecular structures. For example, it is difficult to study complex mixtures of fatty acids of different chain lengths by NMR spectroscopy, because the signals from the different molecular structures are overlaid in the resulting spectrum. Although ever more complex NMR-based approaches have been devised to aid separation and unique identification of given molecular structures (such as 2D-TOCSY or various heteronuclear NMR approaches), these approaches are resource intensive and may still be less informative than mass spectrometry for certain molecular classes.

2.3 Chromatography and Mass Spectroscopy (LC-MS and GC-MS)

Chromatography followed by mass spectrometry is the other major analytical tool that has been used to generate metabonomic profiles. Both liquid chromatography (LC) and gas chromatography (GC) have been used. GC may offer superior resolving power, at least for certain classes of molecules, but it is limited by the lack of volatility of many metabolites. Although this can be overcome to some extent by covalent modification of the sample prior to chromatography, a range of important metabolites still fail to enter the gas phase and hence do not contribute to the resulting metabonomic profile. As a result, the effective size limit for GC-MS is about 700 Da, whereas much larger molecules can contribute to the NMR-, and to some extent, to the LC-MS-derived profiles.

The major advantage of GC-MS and LC-MS is the sensitivity of the technique, which can detect component compounds down to the likely limit of biological relevance (and, indeed, may be too sensitive in some cases, populating the metabolic dataset with large numbers of minor contaminants of the sample, which are not the products of mammalian metabolism). As a result, components such as vitamins and signaling intermediates, which were invisible to the NMR spectrometer, now contribute to the GC-MS-derived metabonomic profile.

The combination of chromatography and mass spectroscopy (MS) can also allow unambiguous assignment of structure to the components of the biological sample in a way that is difficult, though not impossible, with NMR spectroscopy. Comparison of fragmentation patterns with databases, allows deconvolution of peaks with overlapping retention times, and as many as 1000 different molecular components to be unambiguously identified from a single complex biological fluid. This is particularly true when comparing different members of the same homologous series (such as fatty acids of

different chain lengths). However, LC-MS and GC-MS usually poorly resolve structural isomers that are readily distinguished by NMR spectroscopy. It is also important to note that, unlike NMR spectroscopy, mass spectrometry cannot be used to obtain the structure of unknown components contributing the profile – if the fragmentation pattern is not among the database of ''known'' metabolites, no assignment is possible.

Furthermore, MS-based detection is inherently biased: some molecules do not ionize, or only ionize poorly, under any given set of conditions and as a result are either completely invisible or detected only weakly. Consequently, an MS-derived metabonomic profile may be highly biased, and little information can be gained from the relative signals due to different components in the biological sample. However, it is still possible to compare the levels of the same component across different samples in a reliable and reproducible way. Again, variations in the implementation of the technique (such as using both positive and negative ionization modes and varying the cone voltage) can alleviate, though not eliminate, this problem at the cost of increased resource implication.

2.4 Nanosensors

More recently, it has become clear that various designs of the nanosensor can also be applied to generating metabonomic profiles. Nanosensors coated with various hydrophobic coatings adsorb a wide range of metabolites differentially, allowing a profile to be generated that is related in a complex fashion to the molecular composition of the fluid under study. Although to date, no metabonomic profile that has been generated using nanosensor technology has been reported in the scientific literature, it seems likely that such datasets will appear imminently.

Nanosensor-derived metabonomic profiles will presumably have the advantage of low cost (being rapid and high throughput, and not requiring capital intensive reading equipment). However, the complex nature of the relationship between the resulting data vector and the molecular composition of the sample will likely preclude any straightforward listing of the component molecules or estimation of the relative amounts. Such a profile will be neither substantially complete nor unbiased.

Yet, at least for clinical diagnostic applications, such profiles are not without utility. Although it may be difficult (or indeed impossible) to understand precisely which molecular components contribute to the systematic difference in profiles between two groups of interest (such as diseased individuals versus healthy individuals), nevertheless, the very presence of a systematic difference may be diagnostically useful.

2.5 Other Approaches

Many other analytical chemistry techniques can be applied to complex biological fluids to generate a metabonomic profile, although (like nanosensors) such profiles cannot usually be translated into a list of molecular components with associated relative concentrations. For example, infrared spectroscopy provides a low-cost approach to generating a metabonomic profile, which may be useful in some circumstances. It may be possible, depending on the nature of the biological

sample and the particular molecular components of interest, to generate a limited list of molecular components from an infrared spectrum.

Ultimately, however, the utility of a metabonomic profile obtained with any given technique depends on the application. Many studies (particularly aimed at diagnostic applications) may not require the immense structural detail that can be generated using the resource-intensive NMR and MS techniques. Equally, attempts to identify biomarkers associated with a particular phenotype will require a metabonomic profile that more nearly corresponds to the theoretical "ideal" profile. There remains, therefore, a considerable amount of trial and error in the selection of the analytical toolkit best suited to answering a particular question.

3
Methods for Interpreting Metabolic Profiles

3.1
The Problems of Interrogating Very Large Datasets

Much of the power of the metabonomic approach stems from the generation of very large datasets, which are essentially unselected in terms of the contributing components (that is, there was no preexisting hypothesis governing the selection of variables to be measured). As a consequence, special techniques are required to handle the resulting datasets, since extracting meaningful conclusions from datasets with millions of datapoints is a daunting exercise.

The basic aim in interpreting a metabonomic dataset is no different from any conventional multivariate analysis. The value of a dependent (or *Y*-) variable is estimated from a collection of measured *X*-variables (Fig. 3). For example, one might wish to estimate blood pressure from a range of physiological (or metabonomic) measures. While the *Y*-variable to be estimated (or "modeled") may be continuous (like blood pressure), it may equally be a discrete classification variable (which divides each observation into two or more groups, such as the presence and absence of a particular disease).

The problems associated with analyzing very large datasets are basically twofold. Firstly, a metabonomic profile might typically be composed of thousands of datapoints for each individual, yet such a profile may only have been generated from a relatively small number of individuals (tens or at most hundreds per group). Such datasets are typically described as "short and fat," having many more variables than observations (illustrated in Fig. 3). Analyzing such short and fat datasets using conventional multivariate statistics is dangerous, because it becomes increasingly likely that you can construct a model that correctly describes the phenotypic classification of the individuals by chance alone as the number of variables exceeds the number of observations.

Secondly, many of the variables that compose the metabolic profile may be highly correlated with each other. This is a particular problem with spectroscopic data, where neighboring variables are integrals of a continuous spectrum, forcing a relationship between nearby spectral regions. Conventional statistical approaches to multivariate analysis assume that all the predictor variables are independent, and the widespread colinearity of metabonomic profile variables renders the conventional multivariate models error prone.

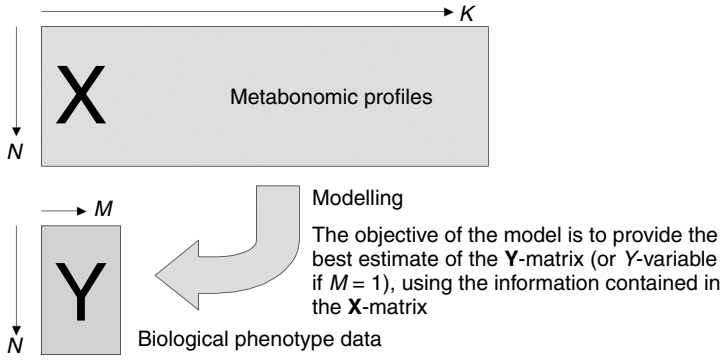

Fig. 3 The principle of regression modeling. The principle of regression modeling is to predict a range of features about a complex system (such as a biological organism) from a collection of unrelated measurements. In metabonomics, this consists of predicting phenotype or behavior (a **Y**-matrix of *M* features of *N* individuals) from a metabonomic dataset consisting of *K* measured metabolic variables from the same *N* individuals. Unlike conventional regression modeling, metabonomics datasets are typically ''short and fat'' with many more measurements per observation than individuals ($K \gg N$). The **Y**-matrix to be predicted might be a single continuous variable ($M = 1$), or even a categorical description (diseased or healthy, for example), whereupon the modeling is usually termed *a discriminant analysis*. Figure reproduced with permission from www.graingerlab.org.

3.2 Conventional Statistical Approaches (LDA)

As noted above, a conventional multivariate model (in which a dependent variable (Y) is predicted from a matrix of independent variables) is based on the assumption that the number of X-variables is less than the number of observations (n), and that all the X-variables are uncorrelated. In most experiments, a series of raw metabonomic profiles violates both assumptions and as a general rule conventional multivariate statistics should not be applied to metabonomic datasets.

However, various preprocessing steps can convert the raw metabonomic dataset into a form amenable to conventional statistical analysis. For example, there exists a range of prefilter algorithms (see Sect. 3.6 below) that allow the number of X-variables to be substantially reduced, for example, by retaining only those X-variables that are most significantly correlated with the dependent variable Y. Such variable-selection algorithms may also incorporate rules to eliminate intercorrelated X-variables. After variable selection, the resulting dataset can be analyzed by conventional multivariate statistics, such as linear discriminant analysis (LDA).

Although such variable-selection approaches circumvent the limitations of conventional multivariate statistics, they can limit the power of the analysis to identify associations between the **X**-matrix and the dependent Y-variable, since much of the information in the **X**-matrix has been discarded. In general, therefore, the other approaches below (which have been developed specifically for the analysis of very

large, or megavariate, datasets) tend to be more powerful.

3.3 Projection Methods (PCA and PLS)

Instead of selecting a subset of variables from the **X**-matrix, it is possible instead to combine the *X*-variables in linear combinations to generate a smaller number of composite variables. This approach (termed *projection*) deals effectively with both of the limitations to megavariate analysis: variable number is reduced, and intercorrelation is minimized.

The principle of projection is illustrated in Fig. 4. Here, a complex 3-dimensional object is represented by a simpler 2-dimensional shadow. In the left panel, the axis of projection is chosen such that the 2-D shadow poorly retains the information of the original object, whereas in the right panel the optimum projection is chosen, which retains most of the information encoded in the original 3-D object. The mathematical algorithms that underlie projection methods such as principal component analysis (PCA) or projection to latent structures using partial least squares (PLS) work by selecting the best projections. A major difference between PCA and PLS is that PLS is a supervised method, which means that the dependent *Y*-variable is used in the process of locating the best projection of the data. Application of these algorithms to high-dimensional datasets (such as metabonomic profiles) can yield just a handful of composite variables (principal components), which are simple linear combinations of the original matrix of *X*-variables. Unlike simpler variable-selection techniques, projection retains as much as possible of the information in the original **X**-matrix, while reducing the dimensionality of the dataset to manageable proportions.

One potential problem with projection methods is overfitting the model. With a sufficiently large number of *X*-variables, it will always be possible to generate combinations of the *X*-variables that predict the dependent variable very well. As a result, it is essential to include a robust external model validation step in the analysis protocol. One example of a useful validation step is scrambling the dependent variable and demonstrating that the **X**-matrix predicts the real dependent variable better than the scrambled variable. An alternative approach is to use the generated model to make predictions about an external dataset not used during model generation. Exhaustive validation is essential for models built using supervised techniques, such as PLS, where the dependent variable was

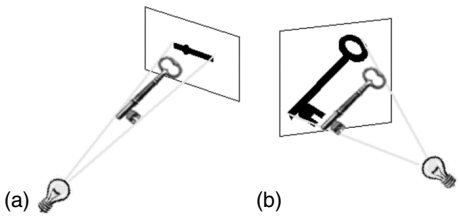

Fig. 4 The principle of projection. A complex high-dimensional object can be represented by a simpler, lower-dimensional model by projection. This is illustrated by the 2-D shadow of 3-D object such as key. However, depending on the particular projection selected, the 2-D model may be poorly representative of the original 3-D object (panel A). The aim of projection modeling tools such as PCA or PLS is to select the optimum lower-dimensional representation of the original complex object (right panel). Figure reproduced with permission from www.graingerlab.org.

used during the construction of the principal components.

3.4 Genetic Computing

Another approach to building models that optimally describe a dependent *Y*-variable from a very large **X**-matrix is to generate a large pool of random models (that are linear or nonlinear combinations of the *X*-variables), a few of which will be acceptable but most of which will be poor, and then apply an evolutionary algorithm to recombine the models and select for improved description of the dependent variable. The evolution is continued through a number of generations, and models can emerge that explain a good proportion of the variation in the *Y*-variable.

This approach can offer a number of advantages over projection-based methods. In particular, the ability to easily include rules that combine *X*-variables in nonlinear ways can be useful when modeling biological systems that have inherent nonlinearities. Another advantage of the evolved models is that they are generally easier to interpret than the optimized projection models. Combinations of simple rules can be more intuitively obvious than lists of principal components

One disadvantage of genetic computing approaches, however, can be ''premature convergence'' whereby the pool of models undergoing evolution rapidly converges on a local maximum of fitness and poorly explores the entire model space. Improvements are continually being made to the basic genetic computing algorithms, and recent advances such as multiobjective fitness functions can alleviate the premature convergence problem.

3.5 Other Approaches

The explosion of high data density analytical techniques in genomics and proteomics, as well as metabonomics, has stimulated the development and refinement of a wide range of other bioinformatics tools to assist in the interpretation of very large datasets. Hierarchical cluster analysis (HCA) is particularly popular for the analysis of gene expression datasets, but it is considerably less powerful than both projection methods and genetic computing algorithms, while offering few advantages. As a result, HCA has not been extensively used to interpret metabonomic datasets.

Another approach, which has been used in metabonomics, is neural network analysis. The neural net is set up with the *X*-variables as inputs and the dependent *Y*-variable as output, with a network of nodes in between. The mathematical function applied at each node is then iteratively varied during a learning phase to optimize the successful prediction of the *Y*-variable. Neural nets can be useful for performing discriminant analysis (that is, classifying observations into two or three groups on the basis of their **X**-matrix values) and hence used for clinical diagnostic purposes, but the models they generate are inherently difficult to interpret, because a large number of structurally different models can yield almost identical predictive power.

3.6 Data Preprocessing and Data Filters

Model building with very large datasets may be best performed in two or even more steps. Rather than directly applying one or more of the model building tools

described above to the raw **X**-matrix, it may be more powerful to perform a preprocessing step first. In fact, as with the model building tools themselves, the object of most of the preprocessing steps is to reduce dimensionality. However, empirical observation suggests that more powerful models can be generated if the dimensionality is reduced in stepwise fashion, perhaps using more than one tool, rather than in a single leap.

Consequently, the most straightforward preprocessing step would be to apply one of the model building tools twice in succession. For example, with projection methods, it is possible (and sometimes useful) to perform an initial principal component analysis reducing the dimensionality of the raw **X**-matrix from, say 8096 variables to tens or hundreds of principal components, and then treat the resulting matrix of principal components as the **X**-matrix for a second round of PCA to reduce the dimensionality down to two or three components, which can be more readily interpreted. The extent to which such hierarchical PCA improves the model compared with application of a single round of PCA depends very much on the particular data structure.

When dealing with metabonomic profiles derived from continuous spectra (such as an NMR spectrum), the neighboring variables (derived from integration of contiguous regions of the spectrum) can be very highly correlated. The shorter the interval along the abscissa that composes each variable, the higher is the degree of local intercorrelation. This can be reduced by integrating over wider intervals (or by averaging neighboring integrals, a process termed *binning*). This effectively reduces the dimensionality of the spectral data, while retaining much of the important information encoded in the spectrum.

Binning relies on the fact that the most highly intercorrelated variables are related to each other by their position along the abscissa in linear fashion. This may not be true for all types of data; in some cases, the most intercorrelated variables may fall cyclically (for example, in the untransformed free-induction decay signal from an NMR spectrometer). For analyzing such datasets, wavelet transformation, rather than binning, will provide the most efficient dimensionality reduction prior to model construction.

Finally, noise filters can be applied to the dataset prior to model building, and for NMR-derived metabonomic datasets this has been shown to significantly improve the performance of projection-based modeling tools. Noise filters (such as orthogonal signal correction; OSC) aim to remove variation in the **X**-matrix, which is uncorrelated with the dependent *Y*-variable, simplifying the dataset for subsequent modeling. It is important to remember that noise filters like OSC are therefore supervised methods, and exhaustive validation of models built on the filtered dataset will be required irrespective of the nature of the modeling tool subsequently used.

4
Applications of Metabonomics

4.1
Probing Normal Human Metabolism

The most obvious application for metabonomics is to aid understanding of normal human metabolism. It is possible to build up a detailed picture of metabolic pathways by taking a time series of metabolic profiles and constructing quantitative models describing the metabolic flux through

various pathways. Although the properties of key metabolic pathways (such as glycolysis or tricarboxylic acid cycle) have been extensively investigated for years, metabonomics can still reveal important aspects of metabolism, which had previously gone unnoticed. In particular, the interaction between endogenous metabolic pathways and the products of symbiotic bacterial metabolism have been extensively investigated using metabonomics.

4.2 Probing the Pathophysiology of Human Disease

Metabonomics is particularly powerful for analyzing the metabolic changes associated with the development of a particular disease state. By comparing metabonomic profiles taken from diseased individuals with profiles from healthy control individuals, it is possible to identify the systematic differences associated with the presence of the disease. What remains challenging, however, is determining which (if any) of these metabolic changes are causes of the disease pathology, as opposed to direct or even indirect consequences of the disease progression.

For example, metabonomics has been used to investigate the metabolic changes associated with the development of osteoporosis. Serum samples from women with low bone mineral density and from healthy control women were subjected to ^{1}H-NMR analysis, and the resultant profiles compared using the projection method PLS-DA following an OSC prefilter. One of the most important metabolites responsible for the separation of the diseased and healthy individuals was the amino acid proline. Women with pathological low bone mineral density had lower levels of proline in their serum; an observation that was later confirmed used conventional biochemical assays. While this does not prove that low serum proline is responsible for the lower bone density, it is a plausible hypothesis for further investigation: proline is a key constituent of the collagen component of bone matrix, and inadequate proline supplies for collagen biosynthesis could represent an entirely novel pathophysiological mechanism in the development of osteoporosis.

4.3 Investigating and Validating Animal Models of Disease

Animal models of human diseases are important tools in scientific and pharmaceutical research and development. The development of genetic manipulation techniques has allowed good models of many monogenic disorders (such as muscular dystrophy) to be developed, confident in the knowledge that the underlying cause of the disease is similar in the animal as in man. In contrast, for complex polygenic disorders such as atherosclerosis or Alzheimer's disease, it is unclear to what extent any given animal model mimics the molecular mechanisms underlying disease susceptibility in man, even if the phenotype of the animals faithfully mirrors the human disease.

Metabonomics offers an opportunity to address this question. Metabolic profiles of animal models can be compared directly with the profiles from human sufferers, allowing a comparison of the physiological perturbations accompanying disease development in the two species. Perhaps more powerfully, it should be possible to track the metabolic trajectory of both animals and humans as the disease progresses, with similar trajectories increasing the confidence in the likely validity

of the model. Studies of this nature are currently underway for a range of animal models of disease, but to date none have been published.

4.4 Clinical Diagnosis of Disease

In many senses, using metabonomics to perform clinical diagnoses is more straightforward than its application to biomarker identification and pathophysiological analysis. Providing a clinically useful diagnosis of a disease on the basis of a serum sample only requires the identification of a robust metabolic signature that always accompanies the disease and is rarely, if ever, present in healthy control individuals. It is not necessary to be able to identify any of the molecular components contributing to the disease-associated metabolic signature.

For example, projection analysis using PLS-DA of ^{1}H-NMR-derived metabonomic profiles of serum samples from individuals with coronary artery disease and healthy control individuals demonstrated clear separation of the two groups. After application of the OSC prefilter, it was possible to predict the disease status of individuals with at least 90% sensitivity and specificity. This metabonomics diagnostic test outperforms all existing noninvasive tests for coronary heart disease by a considerable margin. Although such a test is potentially useful in the clinic (since, at present, the gold-standard diagnostic test for heart disease is an invasive angiography procedure that is expensive and carries a small risk to the patient), it does not readily identify the particular molecular species responsible for separating the two groups. Much of the discriminatory power of the test falls in the region of the NMR spectrum due to lipid components (unsurprisingly, given our knowledge of the mechanisms underlying heart disease) but NMR poorly resolves these closely related lipid structures and considerable further work will be required before the precise molecular basis for the success of the test is known.

4.5 Selection of Subjects for Clinical Trials

At present, a substantial impediment to the testing of new therapeutics for certain diseases is the ability to identify potential sufferers ahead of time. For example, to test a drug proposed to reduce the incidence of myocardial infarction requires the recruitment of subjects at high risk of suffering a myocardial infarction during the trial. Unfortunately, current methods of identifying such subjects are poor, and trials of this nature can require the study of thousands of individuals for three years or more to accumulate sufficient myocardial infarction events for the impact of the drug to be detectable. Although metabonomics-based diagnostics have yet to be used in such applications, it is likely that their widespread adoption will rapidly follow the first successful demonstration of such use.

4.6 Monitoring Efficacy of Therapeutic Interventions

One of the most exciting applications of megavariate diagnostics, whether based on metabonomic, genomic, or proteomic profiles, is the prospect of personalized therapeutic interventions. At present, many drugs are used on broad swathes of the population (for example, statins to lower circulating LDL cholesterol) without any clear indication as to whether they are equally effective in all individuals. Pilot

studies already indicate that it is possible to predict the response of an individual to statin therapy from their metabonomic profile measured prior to beginning therapy. Such ''pharmacometabonomics'' could be extended to optimize the dose and delivery route of a wide range of drugs for each individual, with likely improvements in the efficacy of treatment.

4.7 Toxicology

Perhaps the most mature application of metabonomics is the application of metabolic profiling to toxicology. By studying the metabolic response to a range of model toxins (with organ-specific toxicity), it has been possible to identify metabolic signatures associated with damage to a particular organ. Extensive studies following the metabolic trajectory of animals treated with these model toxins have been published, although commercial considerations mean that few studies of clinically relevant pharmaceutical compositions have reached the public domain.

While metabonomics may help improve the predictivity of animal toxicology studies (which are notoriously difficult to interpret using conventional physiological and histological end points), perhaps the most exciting possibility is the use of metabonomics to perform early stage toxicology directly in man. Because of the sensitivity of metabonomics to detect minute perturbations in the metabolic signature, it may be possible to get an indication of the mode of toxicity of novel chemical entities given in man at doses well below those at which any irreversible damage might occur. The ability to perform meaningful toxicology in man should improve the safety of our medications, and at the same time reduce the number of promising pharmaceutical compounds dropped at a relatively late stage in development because of adverse, and possibly species-specific, side effects observed in the animal models currently used for toxicology.

4.8 Predicting Future Disease Risk

If a metabonomic profile can be used to diagnose the presence of an existing disease (such as coronary heart disease or osteoporosis), there is no reason in principle why it cannot be used to predict future disease susceptibility in the same way that genomic profiles are currently being used. To provide a useful indicator of future disease risk, there must be a component of the dynamic metabolic profile that is temporarily stable on a timescale of years and that is variable between individuals. We have already shown that such a stable interperson variance component exists in NMR-derived metabonomic profiles, and it will be interesting to see to what extent this stable element of the metabolic profile predicts the risk of a range of important diseases.

5 Future Prospects and Challenges

The discipline of metabonomics is expanding rapidly. Although selection from among the many combinations of analytical chemistry and mathematical modeling approaches for any given applications remains empirical, nevertheless, the number of metabonomics studies is growing quickly. Successful examples of the use of metabonomics to answer a broad array of scientific and practical questions are now plentiful in the literature (see

Sect. 4), and application to an ever-broader array of problems seems inevitable. Indeed, it is hard to think of a problem that would not be better addressed by a combination of high data density ''omics'' approaches than by hypothesis-driven reductionist experimentation. It seems that only a deep-seated unease among much of the scientific community for such exploratory, holistic approaches relegates the typical metabonomics experiment to ''second-class status'' and hampers an even more rapid expansion.

One of the challenges facing metabonomics over the coming years, therefore, is to better understand which of the many experimental approaches is optimum for a given use. Doing so will likely require the careful analysis of a single sample set by multiple different analytical chemistry techniques, and then each resulting dataset be interpreted using a range of mathematical modeling tools. In this way, the comparative power of the different approaches will begin to emerge.

Another important challenge is the integration of metabonomics datasets with the large profiles generated by other high data density techniques such as genomics or proteomics. In a sense, the division of ''omics'' science along the lines of the analytical techniques needed to make the measurements is entirely arbitrary. Ultimately, it should prove powerful to combine the profiles obtained from multiple different measurement approaches (whether gene expression, protein levels, or metabolite profiles) into a single ''multi-omics descriptor.'' While there remains considerable debate as to exactly how this amalgamation should be performed, it is widely acknowledged that such a system-wide profile is likely to prove more powerful than a metabonomic or genomic profile alone for many applications. Only such a system-wide profile can allow a complete understanding of such a complex system as a biological organism.

See also Adipocytes.

Bibliography

Books and Reviews

Breiman, L. (2001) Statistical modelling: the two cultures, *Stat. Sci.* **16**(3), 199–231.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S. *Multi-and Megavariate Data Analysis: Principles and Applications*, Umetrics Academy, Umea, Sweden, pp 1–525.

German, J.B., Roberts, M.A., Watkins, S.M. (2003) Personal metabolomics as a next generation nutritional assessment, *J. Nutr.* **133**(12), 4260–4266.

Holmes, E., Antti, H. (2002) Chemometric contributions to the evolution of metabonomics: mathematical solutions to characterising and interpreting complex biological NMR spectra, *The Analyst* **127**(12), 1549–1557.

Kell, D.B. (2004) Metabolomics and systems biology: making sense of the soup, *Curr. Opin. Microbiol.* **7**(3), 296–307.

Kell, D.B., Oliver, S.G. (2004) Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era, *BioEssays* **26**(1), 99–105.

Lindon, J.C., Holmes, E., Bollard, M.E., Stanley, E.G., Nicholson, J.K. (2004) Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis, *Biomarkers* **9**(1), 1–31.

Moolenaar, S.H., Engelke, U.F., Wevers, R.A. (2003) Proton nuclear magnetic resonance spectroscopy of body fluids in the field of inborn errors of metabolism, *Ann. Clin. Biochem.* **40**(Pt 1), 16–24.

Nicholson, J.K., Wilson, I.D. (2003) Understanding 'global' systems biology: metabonomics and the continuum of metabolism, *Nat. Rev. Drug Discov.* **2**(8), 668–676.

Weckwerth, W. (2003) Metabolomics in systems biology, *Annu. Rev. Plant Biol.* **54**, 669–689.

Primary Literature

Beckwith-Hall, B.M., Brindle, J.T., Barton, R.H., Coen, M., Holmes, E., Nicholson, J.K., Antti, H. (2002) Application of orthogonal signal correction to minimise the effects of physical and biological variation in high resolution 1H NMR spectra of biofluids, *The Analyst* **127**(10), 1283–1288.

Blake, G.J., Otvos, J.D., Rifai, N., Ridker, P.M. (2002) Low-density lipoprotein particle concentration and size as determined by nuclear magnetic resonance spectroscopy as predictors of cardiovascular disease in women, *Circulation* **106**(15), 1930–1937.

Boersma, M.G., Solyanikova, I.P., Van Berkel, W.J., Vervoort, J., Golovleva, L.A., Rietjens, I.M. (2001) 19F NMR metabolomics for the elucidation of microbial degradation pathways of fluorophenols, *J. Ind. Microbiol. Biotechnol.* **26**(1–2), 22–34.

Brindle, J.T., Nicholson, J.K., Schofield, P.M., Grainger, D.J., Holmes, E. (2003) Application of chemometrics to 1H NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension, *The Analyst* **128**(1), 32–36.

Brindle, J.T., Antti, H., Holmes, E., Tranter, G., Nicholson, J.K., Bethell, H.W., Clarke, S., Schofield, P.M., McKilligin, E., Mosedale, D.E., Grainger, D.J. (2002) Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics, *Nat. Med.* **8**(12), 1439–1444; Epub 2002 Nov 25. Erratum in: *Nat. Med.* 2003 **9**(4), 477.

Bundy, J.G., Ramlov, H., Holmstrup, M. (2003) Multivariate metabolic profiling using 1H nuclear magnetic resonance spectroscopy of freeze-tolerant and freeze-intolerant earthworms exposed to frost, *Cryo. Letters* **24**(6), 347–358.

Burns, S.P., Woolf, D.A., Leonard, J.V., Iles, R.A. (1992) Investigation of urea cycle enzyme disorders by 1H-NMR spectroscopy, *Clin. Chim. Acta.* **209**(1–2), 47–60.

Bundy, J.G., Spurgeon, D.J., Svendsen, C., Hankard, P.K., Osborn, D., Lindon, J.C., Nicholson, J.K. (2002) Earthworm species of the genus Eisenia can be phenotypically differentiated by metabolic profiling, *FEBS Lett.* **521**(1–3), 115–120.

Choi, Y.H., Kim, H.K., Hazekamp, A., Erkelens, C., Lefeber, A.W., Verpoorte, R. (2004) Metabolomic differentiation of cannabis sativa cultivars using 1H NMR spectroscopy and principal component analysis, *J. Nat. Prod.* **67**(6), 953–957.

Coen, M., Lenz, E.M., Nicholson, J.K., Wilson, I.D., Pognan, F., Lindon, J.C. (2003) An integrated metabonomic investigation of acetaminophen toxicity in the mouse using NMR spectroscopy, *Chem. Res. Toxicol.* **16**(3), 295–303.

Eads, C.D., Furnish, C.M., Noda, I., Juhlin, K.D., Cooper, D.A., Morrall, S.W. (2004) Molecular factor analysis applied to collections of NMR spectra, *Anal. Chem.* **76**(7), 1982–1990.

Engelke, U.F., Liebrand-van Sambeek, M.L., de Jong, J.G., Leroy, J.G., Morava, E., Smeitink, J.A., Wevers, R.A. (2004) N-acetylated metabolites in urine: proton nuclear magnetic resonance spectroscopic study on patients with inborn errors of metabolism, *Clin. Chem.* **50**(1), 58–66.

Freedman, D.S., Otvos, J.D., Jeyarajah, E.J., Shalaurova, I., Cupples, L.A., Parise, H., D'Agostino, R.B., Wilson, P.W., Schaefer, E.J. (2004) Sex and age differences in lipoprotein subclasses measured by nuclear magnetic resonance spectroscopy: the Framingham study, *Clin. Chem.* **50**(7), 1189–1200.

Gavaghan, C.L., Holmes, E., Lenz, E., Wilson, I.D., Nicholson, J.K. (2000) An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse, *FEBS Lett.* **484**(3), 169–174.

Gavaghan, C.L., Nicholson, J.K., Connor, S.C., Wilson, I.D., Wright, B., Holmes, E. (2001) Directly coupled high-performance liquid chromatography and nuclear magnetic resonance spectroscopic with chemometric studies on metabolic variation in Sprague-Dawley rats, *Anal. Biochem.* **291**(2), 245–252.

Griffin, J.L. (2004) Metabolic profiles to define the genome: can we hear the phenotypes? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**(1446), 857–871.

Griffin, J.L., Cemal, C.K., Pook, M.A. (2004) Defining a metabolic phenotype in the brain of a transgenic mouse model of spinocerebellar ataxia 3, *Physiol. Genomics* **16**(3), 334–340.

Griffin, J.L., Troke, J., Walker, L.A., Shore, R.F., Lindon, J.C., Nicholson, J.K. (2000) The biochemical profile of rat testicular tissue as measured by magic angle spinning 1H NMR spectroscopy, *FEBS Lett.* **486**(3), 225–229.

Griffin, J.L., Williams, H.J., Sang, E., Clarke, K., Rae, C., Nicholson, J.K. (2001) Metabolic profiling of genetic disorders: a multitissue (1)H nuclear magnetic resonance spectroscopic and pattern recognition study into dystrophic tissue, *Anal. Biochem.* **293**(1), 16–21.

Hammad, S.M., Powell-Braxton, L., Otvos, J.D., Eldridge, L., Won, W., Lyons, T.J. (2003) Lipoprotein subclass profiles of hyperlipidemic diabetic mice measured by nuclear magnetic resonance spectroscopy, *Metabolism* **52**(7), 916–921.

Harrigan, G.G., LaPlante, R.H., Cosma, G.N., Cockerell, G., Goodacre, R., Maddox, J.F., Luyendyk, J.P., Ganey, P.E., Roth, R.A. (2004) Application of high-throughput Fourier-transform infrared spectroscopy in toxicology studies: contribution to a study on the development of an animal model for idiosyncratic toxicity, *Toxicol. Lett.* **146**(3), 197–205.

Hirabayashi, Y., Matsumoto, Y., Matsumoto, M., Toida, T., Iida, N., Matsubara, T., Kanzaki, T., Yokota, M., Ishizuka, I. (1990) Isolation and characterization of major urinary amino acid O-glycosides and a dipeptide O-glycoside from a new lysosomal storage disorder (Kanzaki disease). Excessive excretion of serine-and threonine-linked glycan in the patient urine, *J. Biol. Chem.* **265**(3), 1693–1701.

Holmes, E., Foxall, P.J., Spraul, M., Farrant, R.D., Nicholson, J.K., Lindon, J.C. (1997) 750 MHz 1H NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease, *J. Pharm. Biomed. Anal.* **15**(11), 1647–1659.

Jonsson, P., Gullberg, J., Nordstrom, A., Kusano, M., Kowalczyk, M., Sjostrom, M., Moritz, T. (2004) A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS, *Anal. Chem.* **76**(6), 1738–1745.

Joshi, L., Van Eck, J.M., Mayo, K., Di Silvestro, R., Blake Nieto, M.E., Ganapathi, T., Haridas, V., Gutterman, J.U., Arntzen, C.J. (2002) Metabolomics of plant saponins: bioprospecting triterpene glycoside diversity with respect to mammalian cell targets, *OMICS* **6**(3), 235–246.

Kell, D.B. (2002) Metabolomics and machine learning: explanatory analysis of complex metabolome data using genetic programming to produce simple, robust rules, *Mol. Biol. Rep.* **29**(1–2), 237–241.

Ketchum, R.E., Rithner, C.D., Qiu, D., Kim, Y.S., Williams, R.M., Croteau, R.B. (2003) Taxus metabolomics: methyl jasmonate preferentially induces production of taxoids oxygenated at C-13 in Taxus x media cell cultures, *Phytochemistry* **62**(6), 901–909.

Keun, H.C., Beckonert, O., Griffin, J.L., Richter, C., Moskau, D., Lindon, J.C., Nicholson, J.K. (2002) Cryogenic probe 13 C NMR spectroscopy of urine for metabonomic studies, *Anal. Chem.* **74**(17), 4588–4593.

Keun, H.C., Ebbels, T.M., Bollard, M.E., Beckonert, O., Antti, H., Holmes, E., Lindon, J.C., Nicholson, J.K. (2004) Geometric trajectory analysis of metabolic responses to toxicity can define treatment specific profiles, *Chem. Res. Toxicol.* **17**(5), 579–587.

Khandelwal, P., Beyer, C.E., Lin, Q., Schechter, L.E., Bach, A.C., II. (2004) Studying rat brain neurochemistry using nanoprobe NMR spectroscopy: a metabonomics approach, *Anal. Chem.* **76**(14), 4123–4127.

Kikuchi, J., Shinozaki, K., Hirayama, T. (2004) Stable isotope labeling of Arabidopsis thaliana for an NMR-based Metabolomics approach, *Plant Cell. Physiol.* **45**(8), 1099–1104.

Kleno, T.G., Kiehr, B., Baunsgaard, D., Sidelmann, U.G. (2004) Combination of 'omics' data to investigate the mechanism(s) of hydrazine-induced hepatotoxicity in rats and to identify potential biomarkers, *Biomarkers* **9**(2), 116–138.

Kraus, W.E., Houmard, J.A., Duscha, B.D., Knetzger, K.J., Wharton, M.B., McCartney, J.S., Bales, C.W., Henes, S., Samsa, G.P., Otvos, J.D., Kulkarni, K.R., Slentz, C.A. (2002) Effects of the amount and intensity of exercise on plasma lipoproteins, *N. Engl. J. Med.* **347**(19), 1483–1492.

Lenz, E.M., Bright, J., Wilson, I.D., Morgan, S.R., Nash, A.F. (2003) 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects, *J. Pharm. Biomed. Anal.* **33**(5), 1103–1115.

Li, Z., Lamon-Fava, S., Otvos, J., Lichtenstein, A.H., Velez-Carrasco, W., McNamara, J.R., Ordovas, J.M., Schaefer, E.J. (2004) Fish

consumption shifts lipoprotein subfractions to a less atherogenic pattern in humans, *J. Nutr.* **134**(7), 1724–1728.

Mitchell, S., Holmes, E., Carmichael, P. (2002) Metabonics and medicine: the biochemical oracle, *Biologist (London)* **49**(5), 217–221.

Moolenaar, S.H., Engelke, U.F., Abeling, N.G., Mandel, H., Duran, M., Wevers, R.A. (2001) Prolidase deficiency diagnosed by 1H NMR spectroscopy of urine, *J. Inherit. Metab. Dis.* **24**(8), 843–850.

Moolenaar, S.H., Gohlich-Ratmann, G., Engelke, U.F., Spraul, M., Humpfer, E., Dvortsak, P., Voit, T., Hoffmann, G.F., Brautigam, C., van Kuilenburg, A.B., van Gennip, A., Vreken, P., Wevers, R.A. (2001) Beta-ureidopropionase deficiency: a novel inborn error of metabolism discovered using NMR spectroscopy on urine, *Magn. Reson. Med.* **46**(5), 1014–1017.

Mortishire-Smith, R.J., Skiles, G.L., Lawrence, J.W., Spence, S., Nicholls, A.W., Johnson, B.A., Nicholson, J.K. (2004) Use of metabonomics to identify impaired fatty acid metabolism as the mechanism of a drug-induced toxicity, *Chem. Res. Toxicol.* **17**(2), 165–173.

Nikiforova, V.J., Gakiere, B., Kempa, S., Adamik, M., Willmitzer, L., Hesse, H., Hoefgen, R. (2004) Towards dissecting nutrient metabolism in plants: a systems biology case study on sulphur metabolism, *J. Exp. Bot.* **55**(404), 1861–1870. Epub 2004 Jun 18.

Ohdoi, C., Nyhan, W.L., Kuhara, T. (2003) Chemical diagnosis of Lesch-Nyhan syndrome using gas chromatography-mass spectrometry detection, *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **792**(1), 123–130.

Ohse, M., Matsuo, M., Ishida, A., Kuhara, T. (2002) Screening and diagnosis of beta-ureidopropionase deficiency by gas chromatographic/mass spectrometric analysis of urine, *J. Mass. Spectrom.* **37**(9), 954–962.

Ott, K.H., Aranibar, N., Singh, B., Stockton, G.W. (2003) Metabonomics classifies pathways affected by bioactive compounds, Artificial neural network classification of NMR spectra of plant extracts. *Phytochemistry* **62**(6), 971–985.

Otvos, J.D., Jeyarajah, E.J., Bennett, D.W., Krauss, R.M. (1992) Development of a proton nuclear magnetic resonance spectroscopic method for determining plasma lipoprotein concentrations and subspecies distributions from a single, rapid measurement, *Clin. Chem.* **38**(9), 1632–1638.

Pham-Tuan, H., Kaskavelis, L., Daykin, C.A., Janssen, H.G. (2003) Method development in high-performance liquid chromatography for high-throughput profiling and metabonomic studies of biofluid samples, *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **789**(2), 283–301.

Plumb, R.S., Stumpf, C.L., Gorenstein, M.V., Castro-Perez, J.M., Dear, G.J., Anthony, M., Sweatman, B.C., Connor, S.C., Haselden, J.N. (2002) Metabonomics: the use of electrospray mass spectrometry coupled to reversed-phase liquid chromatography shows potential for the screening of rat urine in drug development, *Rapid Commun. Mass Spectrom.* **16**(20), 1991–1996.

Purohit, P.V., Rocke, D.M., Viant, M.R., Woodruff, D.L. (2004) Discrimination models using variance-stabilizing transformation of metabolomic NMR data, *OMICS* **Summer; 8**(2), 118–130.

Raamsdonk, L.M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M.C., Berden, J.A., Brindle, K.M., Kell, D.B., Rowland, J.J., Westerhoff, H.V., van Dam, K., Oliver, S.G. (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations, *Nat. Biotechnol.* **19**(1), 45–50.

Robertson, D.G., Reily, M.D., Albassam, M., Dethloff, L.A. (2001) Metabonomic assessment of vasculitis in rats, *Cardiovasc. Toxicol.* **1**(1), 7–19.

Sato, S., Soga, T., Nishioka, T., Tomita, M. (2004) Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection, *Plant J.* **40**(1), 151–163.

Slim, R.M., Robertson, D.G., Albassam, M., Reily, M.D., Robosky, L., Dethloff, L.A. (2002) Effect of dexamethasone on the metabonomics profile associated with phosphodiesterase inhibitor-induced vascular lesions in rats, *Toxicol. Appl. Pharmacol.* **183**(2), 108–109.

Soedamah-Muthu, S.S., Chang, Y.F., Otvos, J., Evans, R.W., Orchard, T.J. (2003) Pittsburgh epidemiology of diabetes complications study. Lipoprotein subclass measurements by nuclear magnetic resonance spectroscopy improve the prediction of coronary artery

disease in Type 1 diabetes. A prospective report from the Pittsburgh epidemiology of diabetes complications study, *Diabetologia* **46**(5), 674–682.

Tang, H., Wang, Y., Nicholson, J.K., Lindon, J.C. (2004) Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma, *Anal. Biochem.* **325**(2), 260–272.

Tate, A.R., Foxall, P.J., Holmes, E., Moka, D., Spraul, M., Nicholson, J.K., Lindon, J.C. (2000) Distinction between normal and renal cell carcinoma kidney cortical biopsy samples using pattern recognition of (1)H magic angle spinning (MAS) NMR spectra, *NMR Biomed.* **13**(2), 64–71.

Van, Q.N., Chmurny, G.N., Veenstra, T.D. (2003) The depletion of protein signals in metabonomics analysis with the WET-CPMG pulse sequence, *Biochem. Biophys. Res. Commun.* **301**(4), 952–959.

Verhoeckx, K.C., Bijlsma, S., Jespersen, S., Ramaker, R., Verheij, E.R., Witkamp, R.F., Van Der Greef, J., Rodenburg, R.J. (2004) Characterization of anti-inflammatory compounds using transcriptomics, proteomics, and metabolomics in combination with multivariate data analysis, *Int. Immunopharmacol.* **4**(12), 1499–1514.

Viant, M.R. (2003) Improved methods for the acquisition and interpretation of NMR metabolomic data, *Biochem. Biophys. Res. Commun.* **310**(3), 943–948.

Wang, Y., Bollard, M.E., Keun, H., Antti, H., Beckonert, O., Ebbels, T.M., Lindon, J.C., Holmes, E., Tang, H., Nicholson, J.K. (2003) Spectral editing and pattern recognition methods applied to high-resolution magic-angle spinning 1H nuclear magnetic resonance spectroscopy of liver tissues, *Anal. Biochem.* **323**(1), 26–32.

Watkins, S.M., Reifsnyder, P.R., Pan, H.J., German, J.B., Leiter, E.H. (2002) Lipid metabolome-wide effects of the PPARgamma agonist rosiglitazone, *J. Lipid. Res.* **43**(11), 1809–1817.

Wevers, R.A., Engelke, U., Heerschap, A. (1994) High-resolution 1H-NMR spectroscopy of blood plasma for metabolic studies, *Clin. Chem.* **40**(7 Pt 1), 1245–1250.

Wevers, R.A., Engelke, U.F., Moolenaar, S.H., Brautigam, C., de Jong, J.G., Duran, R., de Abreu, R.A., van Gennip, A.H. (1999) 1H-NMR spectroscopy of body fluids: inborn errors of purine and pyrimidine metabolism, *Clin. Chem.* **45**(4), 539–548.

Xu, J., Chang, V., Joseph, S.B., Trujillo, C., Bassilian, S., Saad, M.F., Lee, W.N., Kurland, I.J. (2004) Peroxisomal proliferator-activated receptor alpha deficiency diminishes insulin-responsiveness of gluconeogenic/glycolytic/pentose gene expression and substrate cycle flux, *Endocrinology* **145**(3), 1087–1095.

Yu, H.H., Ginsburg, G.S., O'Toole, M.L., Otvos, J.D., Douglas, P.S., Rifai, N. (1999) Acute changes in serum lipids and lipoprotein subclasses in triathletes as assessed by proton nuclear magnetic resonance spectroscopy, *Arterioscler. Thromb. Vasc. Biol.* **19**(8), 1945–1949.