Review

The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball?

Julian L. Griffin*

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK

To date most global approaches to functional genomics have centred on genomics, transcriptomics and proteomics. However, since a number of high-profile publications, interest in metabolomics, the global profiling of metabolites in a cell, tissue or organism, has been rapidly increasing. A range of analytical techniques, including ^1H NMR spectroscopy, gas chromatography–mass spectrometry (GC–MS), liquid chromatography–mass spectrometry (LC–MS), Fourier Transform mass spectrometry (FT–MS), high performance liquid chromatography (HPLC) and electrochemical array (EC–array), are required in order to maximize the number of metabolites that can be identified in a matrix. Applications have included phenotyping of yeast, mice and plants, understanding drug toxicity in pharmaceutical drug safety assessment, monitoring tumour treatment regimes and disease diagnosis in human populations. These successes are likely to be built on as other analytical and bioinformatic approaches are developed to fully exploit the information obtained in metabolic profiles. To assist in this process, databases of metabolomic data will be necessary to allow the passage of information between laboratories. In this prospective review, the capabilities of metabolomics in the field of medicine will be assessed in an attempt to predict the impact this ‘Cinderella approach’ will have at the ‘functional genomic ball’.

Keywords: phenotyping; transcriptomics; proteomics; genomics; pattern recognition; bioinformatics

1. INTRODUCTION

The term metabolomics (and the related term metabonomics) was coined at the end of the 1990s to describe the development of approaches which aim to measure all the metabolites that are present within a cell, tissue or organism during a genetic modification or physiological stimulus (Oliver *et al.* 1998; Nicholson *et al.* 1999; figure 1). In this respect, there are obvious parallels with the definitions of transcriptomics and proteomics as tools for measuring global mRNA and protein expression, respectively. The major technologies that have been used for this process include ^1H NMR spectroscopy, gas chromatography–mass spectrometry (GC–MS) and liquid chromatography–mass spectrometry (LC–MS), in conjunction with pattern recognition approaches (Griffin *et al.* 2003; figure 2). However, because of the very diverse range of metabolites found within the cell, in conjunction with the large dynamic range of metabolite concentrations, none of these technologies, even when used in conjunction with one another, can be relied upon to generate a complete metabolomic description of even the simplest organism. To date the upper range of metabolite identification and quantification is of the order of hundreds (Fiehn 2002), which can be compared with the thousands of proteins that can be

analysed by proteomic approaches and the capability of measuring mRNA expression across the complete genome of many organisms. In part for this reason, and in part because of the relatively recent description of metabolomic approaches, interest in metabolomics has lagged behind transcriptomics and proteomics.

However, despite the immense technical challenges associated with analysing a wide range of metabolites of varying concentrations the approaches being developed have several distinct advantages. These techniques are high throughput and are cheap on a per sample basis, making them ideal as rapid screening tools. This has led to the development of a number of screening assays for human populations for diseases such as coronary artery disease (Brindle *et al.* 2002) and diabetes, as well as its use in toxicology as part of the drug safety assessment process (Nicholson *et al.* 2002; Griffin & Bollard 2004). Metabolomics has also proven sensitive enough to a number of subtle genetic modifications including silent mutations in yeast (Raamsdonk *et al.* 2001; Allen *et al.* 2003) and understanding ecotypes in *Arabidopsis* (Fiehn *et al.* 2000). In this prospective review, the capabilities of metabolomics in the field of medicine will be assessed in an attempt to predict the impact this technology will make to the field of functional genomics and systems biology, and in particular how the approach may compete with and complement the more widespread approaches of transcriptomics and

*jlg40@mole.bio.cam.ac.uk

Figure 1. A schematic view of metabolomics. The position of metabolomics is shown in respect to the other omic approaches and the influence of the environment. In addition a scheme is shown for producing pattern recognition models capable of deducing metabolic profiles. The initial phase usually involves the acquisition of a large dataset in terms of both the variables (metabolites) and observations (subjects), commonly using either ^1H NMR spectroscopy or mass spectrometry. However, the use of pattern recognition techniques is an integral part of the approach if the innate metabolic variation associated with different individuals is to be separated from that induced by the insult or physiological stimulus. Following the generation of a database from metabolic responses, this can then be used to build a predictive pattern recognition model that is capable of predicting class membership (e.g. clustering according to the gene deleted).

proteomics. In order to do this, a brief review of the key publications in the field is given.

2. RECENT SUCCESS STORIES IN METABOLOMICS

(a) *Comparing silent phenotypes in yeast*

Yeast was the first eukaryote to be sequenced and cell banks exist for all the ~ 6000 mutants which are viable in this organism (for example, the EU funded project EUROFAN; Oliver *et al.* 1998). While the challenge of defining a complete metabolome for an organism is confounded by the need to define gene–environment interactions for every physiological or pathological stimulus, the existence of such cell banks suggests that it may be possible to phenotype every viable mutant in yeast. This could have a significant impact on human disease through comparison of gene sequences found in yeast or correlating phenotype and metabolic profiles found in both yeast and man.

To carry out a comprehensive genomic screen of all the viable mutants of an organism a high throughput rapid phenotyping tool is required. The standard way to phenotype yeast strains is to examine how rapidly a strain grows on given substrate mixtures; for example comparing normal media with carbon and nitrogen limited media. However, if the mutation does not alter the rate of growth it is said to be a silent mutation, and it is found that even mutations involved in energy metabolism may conspire with the network of

metabolic pathways in yeast to produce a silent phenotype. These metabolic perturbations should be detectable by metabolomics and both NMR and mass spectrometry have been used to distinguish various silent phenotypes in yeast (Raamsdonk *et al.* 2001; Allen *et al.* 2003; Steuer *et al.* 2003). Furthermore, the detected metabolic profiles can be used to cluster genes of a similar function according to the perturbations the genes induce.

In a proof of concept paper Raamsdonk *et al.* (2001) used ^1H NMR spectroscopy to study the metabolic changes induced in the different yeast mutants including two mutants of 6-phosphofructo-2-kinase. Not only did these mutants cluster together following principal components analysis (PCA) and discriminant function analysis, but also mutants of oxidative phosphorylation clustered together. This phenotyping approach can also be applied to the media the yeast is grown in, with this variation to the approach being referred to as ‘metabolic footprinting’ (Allen *et al.* 2003). Indeed, metabolic footprinting has several advantages over conventional analysis of cell extracts, including easy sampling without destroying the cells and the analysis of excretory metabolites which may concentrate in the media to far higher concentrations than would be viable within the cell.

(b) *Genetic modification and environmental interactions in Arabidopsis thaliana*

The challenge of metabolomics in the plant kingdom is potentially larger than that in animal and microbial metabolism. While plant genomes typically contain 20 000–50 000 genes there are currently 50 000 metabolites identified in plants with the total number estimated to rise to $\sim 200\,000$ metabolites (De Luca & St Pierre 2000).

In one of the landmark papers in plant metabolomics, Fiehn *et al.* (2000) used GC–MS to identify and quantify 326 distinct metabolites in *Arabidopsis* leaf extracts. Furthermore, they assigned the chemical structure to half of these metabolites. PCA of the data was able to separate four genotypes representing the interactions between a gene mutation and different wild type backgrounds associated with a given ecotype. Since this paper, the researchers have expanded their approach to include GC–time of flight–MS to detect and characterize ~ 1000 metabolites (Fiehn 2001, 2002). They have also used the huge datasets they have acquired from different mutants to identify metabolites which appear correlated during a given biological intervention, referring to these as ‘metabolic cliques’ (Steuer *et al.* 2004). It is hoped that understanding these metabolic cliques will give a better understanding of how metabolites interact to form a complete metabolic network in an organism and how genetic modifications can cause perturbations in this network. Since these landmark papers a number of researchers have exploited the sensitivity and reproducibility of GC–MS based metabolomics for studying plant metabolism (for example, read Gullberg *et al.* 2004 for a discussion of optimizing the process).

(c) *A high throughput diagnosis of coronary artery disease*

A diagnostic tool for any common human disease must be capable of scaling up throughput, without significant costs per sample, so that the tool could potentially be used to screen a large population. With the advent of flow-probe technology NMR spectroscopy is potentially one such approach in terms of screening for high-concentration metabolite biomarkers. This has allowed a number of research groups to screen for metabolic disorders in urine from humans.

One such study has used pattern recognition approaches to analyse ^1H NMR spectra of blood plasma in order to determine whether patients had coronary artery disease and predict the severity of this disease (Brindle *et al.* 2002). The mathematical analysis of the data involved a combined use of a data filtering procedure called orthogonal signal correction (OSC) followed by PCA or partial least squares-discriminate analysis (PLS-DA). OSC removes variation that is orthogonal to the variation associated with the process of interest, and hence removes variation that is not correlated with a disease, toxic insult or other manipulation being examined in a metabolomic study.

The subsequent pattern recognition models were capable of predicting the occurrence and severity of the disease with over 90% accuracy, reporting the severity as one, two or three vessel disease. This analysis could also be performed on samples from both males and females, and patients on statins, drugs used to treat high cholesterol and blood pressure. However, the current gold standard for detecting coronary artery disease is set by angiography, which is $\sim 99\%$ accurate. Since the publication of this manuscript, the researchers have begun a study entitled MAGE-CAD to use both combined metabolomics and transcriptomics, based on ^1H NMR spectroscopy and mRNA microarrays. It is hoped the increased data from the DNA microarrays will further increase the predictability of the pattern recognition models produced so that a biomarker approach will be as diagnostically robust as angiography. Furthermore, such a ^1H NMR/pattern recognition analysis of blood plasma has also correlated metabolite changes to measured blood pressure (Brindle *et al.* 2003). The possibility of NMR spectroscopy playing a relevant role in diagnosis in hospitals is further augmented by the relative cheapness on a per sample basis to carry out such analyses (typically less than £1), the increased automation and simplification of spectrometer operation, and the introduction of shielded magnets which allow the housing of magnets within much smaller rooms.

(d) *A marker of peroxisomal proliferation in rats and man*

There are three types of peroxisome proliferators-activated receptors (PPARs) referred to as PPAR- α , PPAR- γ and PPAR- δ . Dysfunction of these receptors has been linked to a number of metabolic disorders including type II diabetes and dyslipidaemia. Furthermore, stimulation of these receptors by drugs has produced some of the most effective methods for treating these diseases by stimulating the expression of enzymes involved in β -oxidation and fatty acid

degradation. However, one potential concern of agonists of the PPAR family is that some drugs, as well as changing enzyme expression, have produced large scale peroxisome proliferation which leads to liver dysfunction and also hepatic cancer in rats. This serious potential side effect is further compounded by the fact that peroxisome proliferation can only be monitored by electron microscopy of liver tissue. Thus, unless biomarkers for peroxisome proliferation can be identified, this side effect could not be monitored in humans during clinical trials.

To address this issue Ringeissen *et al.* (2004) examined the urinary metabolites of rats following exposure to two PPAR ligands using a combination of ^1H NMR spectroscopy and PCA. They identified two biomarkers, *N*-methylnicotinamide and *N*-methyl-4-pyridone-3-carboxamide (4PY) both metabolites found in the tryptophan- NAD^+ pathway. Furthermore, the concentrations of both these metabolites were correlated with peroxisome proliferation as measured by electron microscopy during post-mortem analysis of the liver tissue. Real time PCR confirmed transcriptional changes on the tryptophan- NAD^+ pathway, indicating agreement between both metabolomics and transcriptomics in terms of deducing the mechanism of action of these drugs. Since then, a LC-MS based assay has been developed for the two biomarkers in order to monitor potential peroxisome proliferation during the stimulation of the PPAR receptors.

(e) *Monitoring gene therapy in Duchenne muscular dystrophy*

Duchenne muscular dystrophy is an X-linked disorder affecting 1 in 3500 male births (Hoffman *et al.* 1987). The disease is characterized by progressive muscle wasting as well as a non-progressive neurological deficit (Hoffman *et al.* 1987; Emery 1989). This reflects that dystrophin is expressed in three full length isoforms as M-dystrophin in muscle tissue, C-dystrophin in the cortex and P-dystrophin in the cerebellum. While the exact role of dystrophin in the brain is still largely undetermined, in muscle tissue dystrophin provides stability to the sarcolemmar membrane.

The *mdx* mouse is a mouse model of Duchenne muscular dystrophy possessing a gene deletion for dystrophin (Tanabe *et al.* 1986). While the disease in humans is characterized by severe muscle wasting culminating in death within the second decade (Hoffman *et al.* 1987; Emery 1989), in the *mdx* mouse deletion of dystrophin produces a mild pathology, with only a small reduction in longevity in this animal (Tinsley *et al.* 1996), a pronounced hump, and lower muscle regenerative capacity (McIntosh *et al.* 1998a,b).

A number of studies have already demonstrated that *mdx* mice are categorized by both increased concentration of taurine and decreased creatine in the heart and muscle tissue in general (McIntosh *et al.* 1998a,b; Griffin *et al.* 2001). This included one study which simultaneously also identified metabolic fingerprints associated with a failure to express dystrophin in the cortex and cerebellum (Griffin *et al.* 2001). In these studies the concentration of taurine has been correlated with the ability of muscle tissue to regenerate

Figure 2. A comparison of NMR and LC-MS analysis of urinary metabolite profiles. A range of different profiling tools are currently being used for metabolomics, with no one approach providing a complete coverage of the metabolome. These figures show PCA of data derived from a study of rat urine using NMR spectroscopy (a) and LC-MS (b). In these figures each dot represents a sample of rat urine analysed. Samples with similar metabolic profiles will have a similar position in the PCA scores plot, producing the clustering detected. In both models the metabolite profiles could be classified according to gender of the animal or the time the samples were removed. However, in this example more discrete classification was produced using LC-MS. Diagram provided by Mark Hodson and John Haselden of GlaxoSmithKline.

(McIntosh *et al.* 1998a). Using a combination of both ^1H NMR spectroscopy and pattern recognition techniques, Griffin *et al.* (2002) investigated a mouse model based on the *mdx* mouse where the lack of dystrophin in muscle tissue, but not cardiac tissue, was compensated for by the upregulation of utrophin, a dystrophin related protein. Increasing utrophin expression has raised much interest as a potential

therapy for DMD, as muscle wasting is only apparent in sufferers of DMD after utrophin expression decreases shortly after birth. Furthermore, the dystrophic phenotype normally observed in *mdx* mice is absent when muscles overexpress utrophin (Tinsley *et al.* 1998), and utrophin based gene therapy has been suggested as one potential treatment for Duchenne muscular dystrophy (Burton & Davies 2002).

The concentration of taurine in skeletal muscle of *mdx* mice expressing the utrophin transgene ($Tg_{fulllength}/Dmd^{mdx}$) was intermediate between *mdx* and control mice, suggesting that the metabolic profiles determined by 1H NMR spectroscopy could potentially be used to follow gene therapy. One role for the increased taurine in dystrophic tissue could be to compensate for increased intracellular Ca^{2+} in the tissue (Huxtable 1992).

Following on from these studies of DMD, Jones *et al.* (2005; figure 3) have simultaneously modelled the metabolic profiles produced by 1H NMR spectroscopy of extracts of heart tissue from the *mdx* mouse, a mouse model of cardiac hypertrophy (muscle LIM protein knock out mouse) and two mouse models of cardiac arrhythmia. The metabolic profiles demonstrated that strain background is an important component of the global metabolic phenotype of a mouse, providing insight into how a given gene deletion may result in very different responses in diverse populations. Despite these differences associated with strain, multivariate statistics was capable of separating each mouse model from its control strain, demonstrating that metabolic profiles could be generated for each disease. Thus, metabolomics provides a rapid method of phenotyping mouse models of disease.

(f) Understanding apoptosis in tumours

While the use of NMR spectroscopy as a tool for metabolomics is sometimes limited by the sensitivity of the approach, one very important aspect of the approach is that NMR can be applied *in vivo* (referred to as magnetic resonance spectroscopy—MRS). Perhaps the area that this has most been demonstrated is the field of cancer research. Studies using MRS *in vivo* have been able to distinguish a number of tumour types (for example, Tate *et al.* 1996a, 1998) despite *in vivo* spectroscopy being limited by sensitivity and broad metabolite resonances, while *in vitro* studies of tumours and cell lines have also been able to suggest how these tumours differ in terms of biochemical pathways (Florian *et al.* 1995a,b). In addition high resolution magic angle spinning (HRMAS) 1H NMR spectroscopy, an approach capable of producing high-resolution spectra from intact tissue, has also been widely used to follow tumour metabolism (Cheng *et al.* 1996; Tomlins *et al.* 1998; Millis *et al.* 1999; Chen *et al.* 2001; Tate *et al.* 2003).

One of the first applications of metabolomics both *in vivo* and *in vitro* involved the investigation of hypoxia inducible factor (HIF)-1 β in tumour metabolism and growth (Griffiths *et al.* 2002, Griffiths & Stubbs 2003). The expression of HIF-1 β is increased in a number of cancers resulting in the upregulation of proteins in a number of metabolic pathways including glucose transporters, glycolytic pathways and vascular endothelium growth factor, thus increasing the flux through glycolysis in these tumours. Hepatoma cells deficient in HIF-1 β grown as solid tumours in mice have reduced rates of growth, but magnetic resonance imaging indicated no decreased vascularity in HIF-1 β deficient tumours, despite marked metabolic dysfunction in terms of ATP content as measured by ^{31}P NMR spectroscopy.

Figure 3. An orthogonal signal corrected (OSC) PLS-DA analysis of four mouse models of cardiac disease using metabolic profiles from 1H NMR analysis of cardiac tissue extracts. OSC is a data filtering technique that removes variation uncorrelated with class membership. The PLS-DA analysis then clusters spectra from the different mouse models across three axes that represent the most amount of correlated variation in the dataset associated with class membership. Key: open circle, control animals (three different strains); plus, mouse model of Duchenne muscular dystrophy; filled diamond, mouse model of cardiac hypertrophy (Muscle LIM protein knock out mouse), and two models of cardiac arrhythmia: a cardiac sodium channel knock out mouse ($Scn^{-/+}$), filled square, and a model where the closure of the cardiac sodium channel is impaired ($Scn^{\Delta/+}$), filled triangle (taken from Jones *et al.* 2005).

In vitro NMR spectroscopy indicated a significant decrease in phosphocholine, choline, betaine and glycine, and suggested the cause of the reduced ATP content of the mutant cells. Glycine is formed from the glycolytic intermediate 3-phosphoglycerate and is an important source of one-carbon units for the synthesis of nucleotides, via serine. Thus, if HIF-1 β deficient tumours had reduced glycolytic flux, and hence glycine content, this would impair nucleotide synthesis producing a decrease in ATP. As glycine can also be produced via choline and betaine this also explained the other results produced by NMR-based metabolomics.

A combination of *in vivo*, *in vitro* and HRMAS 1H NMR spectroscopy has also been used in conjunction with PCA to examine polyunsaturated fatty acids (PUFA) which accumulate in glioma during gene therapy-induced programmed cell death (PCD) (Griffin *et al.* 2003; Lehtimäki *et al.* 2003; Valonen *et al.* in press). Metabolomics demonstrated that the concentration of 1H NMR detectable PUFAs increased threefold, with pattern recognition identifying $CH=CH$ and $CH=CHCH_2CH=CH$ as being the most significant in monitoring the dynamics of PCD. Furthermore, a range of biophysical data obtained using relaxation and diffusion rate measurements by NMR indicated the

lipids found to increase during apoptosis were most likely in cytoplasmic lipid vesicles. The most likely cause of these ^1H NMR detectable lipids formed during PCD is the breakdown of cell constituents forming cytoplasmic lipid vesicles in dying cells.

One of the problems with interpreting NMR spectra obtained *in vivo* and by HRMAS ^1H NMR is co-resonant metabolites, where a number of chemically related metabolites may have overlapping resonances. This suggests that these studies may benefit from further examination of tissue extracts using GC- and LC-mass spectrometry or LC-NMR capable of separating many chemically similar metabolites prior to analysis. However, Griffin *et al.* (2003) have shown that HRMAS ^1H NMR spectroscopy detects the metabolically active metabolites present in a tissue, as opposed to components of the cell membrane and other compartments isolated from normal metabolism, and conversely techniques relying on tissue extracts tend to sample the entire cell or tissue.

3. HOW MANY METABOLITES DO WE NEED TO MEASURE?

There have been various estimates made of the total number of metabolites found within a cell. Of course, the metabolite complement will be dependent on the exact cell type, with cells from multicellular organisms expected to have a smaller number of metabolites because of differentiation and specialization compared with unicellular organisms. A simple consideration of the metabolites that appear in the metabolic pathways found in mammalian cells estimates the metabolite complement of a mammalian cell to be ~ 650 , although this number does not reflect the diversity of lipid modifications possible. For example, if it is assumed that there are about 40 commonly occurring fatty acids within the cell, then triglycerides, each containing three fatty acids esterified to a central glycerol backbone, could represent a group consisting of $40 \times 40 \times 40$ or 64 000 metabolites. Applying a similar analysis to fatty acids, fatty acid acyl CoA, monoacylglycerides, diacylglycerides and triglycerides this produces a total of 69 000 metabolites (Berger 2004). This simple calculation does not take into consideration the preference of certain types of fatty acids to particular positions on the glycerol backbone, but does demonstrate the potential diversity of metabolites. Furthermore, the dynamic range of metabolites is likely to be of the order of 10^9 with a wide range of polarities. However, do we need to measure all of these metabolites in order to understand metabolism?

High-resolution ^1H NMR spectroscopy is a relatively insensitive technique which only measures the high-concentration metabolites found within a tissue extract, biofluid or intact tissue. Typically, using a simple one-dimensional pulse sequence 30–100 metabolites are observable in urine, 20–30 metabolites in blood plasma or serum and 10–30 metabolites in tissue extracts. Despite this insensitivity, ^1H NMR based metabolomics (often termed metabonomics) has been successfully used to distinguish a range of liver and kidney toxins in the rat (Beckwith-Hall *et al.* 1998; Holmes *et al.* 1998; Mortishire-Smith *et al.* 2004),

a number of cardiac diseases in mouse models (Griffin *et al.* 2002; Jones *et al.* 2005) and silent phenotypes in yeast (Raamsdonk *et al.* 2001). One can begin to understand these conflicting viewpoints when we consider the metabolites that ^1H NMR spectroscopy detects. Many of these high-concentration metabolites are found in a number of metabolic pathways, and in terms of the metabolic network of the cell these metabolites represent central hubs of metabolism (Brindle 2003; Griffin *et al.* 2003). Thus, a perturbation at one point in the metabolic network can rapidly be transferred to other metabolic pathways through these highly connected metabolites. Thus, we would expect the high-concentration metabolites to be highly variable during a perturbation, and hence ^1H NMR spectroscopy sensitive to a wide range of genetic modifications, toxicology insults and physiological stimuli.

However, there is a negative side to the detection of high-concentration metabolites. If these metabolites vary across a wide range of modifications, it then becomes difficult to use these metabolites as unique biomarkers for disease processes or for deducing which pathways have been stimulated during a modification. This can be further confounded by the detection of non-specific effects, which accompany a genetic modification or toxicological insult. For example, Connor *et al.* (2004) have observed that many so-called biomarkers of liver and kidney toxicity, including hippurate, taurine, citrate and 2-oxoglutarate, in fact arise from the effects of food restriction induced in animals that are unwell following a toxic insult. This illustrates the importance of using a pair fed control group in investigative toxicology studies if the metabolic markers identified are to be used reliably to deduce which metabolic pathways are perturbed. It also highlights the sensitivity of metabolomic studies to changes associated with general stress, which should be considered when carrying out toxicology tests.

An alternative to measuring metabolic fingerprints of high-concentration metabolites is to use approaches capable of detecting unique biomarkers. Because of the unique concentrating power of the kidneys, it is possible to detect such biomarkers in the ^1H NMR spectra of the urine, and biomarkers have been postulated for testicular toxicity (Nicholson *et al.* 1989), phospholipidosis (Nicholls *et al.* in press) and peroxisome proliferation (Ringeissen *et al.* 2004). However, in general for subtle genetic modifications or mild drug interventions discrete biomarkers are not detectable among the high-concentration low-molecular weight metabolites detectable by conventional ^1H NMR spectroscopy. An alternative is to use more sensitive techniques such as GC-MS or LC-MS to detect lower concentration metabolites. These metabolites should be far more discriminatory compared with metabolic fingerprints, and from a theoretical basis the majority of metabolites within the cell are involved in a single pathway. However, despite the potential of mass spectrometry based approaches to identify unique biomarkers for disease and toxic insults studies, none have been reported currently in the literature.

Figure 4. Principal component analysis of urinary metabolite profiles (left) readily separated two strains of mouse. Inspection of the loadings plot for these models showed that the two mouse strains were categorized according to differences in the metabolic pathways associated with trimethylamine-oxide metabolism (right). Taken from [Gavaghan *et al.* 2000](#).

4. 'POLY-OMICS': CROSS CORRELATION OF METABOLOMICS AND OTHER '-OMIC' TECHNOLOGIES

(a) Genomics and metabolomics

One of the first uses of metabolomics was as a rapid phenotyping tool for interesting mutants produced in plants, animals and microbes ([Fiehn *et al.* 2000](#); [Gavaghan *et al.* 2000](#); [Raamsdonk *et al.* 2001](#); [Jones *et al.* 2005](#)). This has produced some surprises in terms of our understanding of functional genomics. For example, when [Fiehn *et al.* \(2000\)](#) examined mutants of *Arabidopsis* using GC-MS they found that the dominant effects in the samples they examined were related to 'ecotypes' (the genetic background associated with a particular ecological niche) rather than the mutants *per se*.

Similarly, in mice, [Gavaghan *et al.* \(2000\)](#) found that C57BL10J and Alpk : ApfCD mouse strains of healthy mice could readily be separated by the metabolic profiles obtainable from ^1H NMR spectroscopic analysis of urine ([figure 4](#)). This largely resulted from TCA cycle intermediates and methylamine pathway products, demonstrating that some of the key metabolic intermediates were highly variable across different strains. Furthermore, they were able to postulate that one of the strains may be predisposed to certain metabolic disorders as a result of the metabolic phenotype, or 'metabotype' associated with that mouse strain. The importance of strain background has previously been noted in a number of physiological studies. [Bendall *et al.* \(2002\)](#), investigating vascular responses to nitric oxide (NO) in the murine heart, found distinct strain dependences. Comparing MF1, C57BL/6J and 129SV mouse strains, the hearts from the three strains differed in responses to endothelium-derived vasodilators, exogenous NO, eNOS expression and superoxide production. The authors concluded that as an aspect of many pathologies, including cardiac hypertrophy, hypertension and heart failure, was the generation of reactive oxygen species, the strain background might have a profound effect on the development of the disease

in a given murine model. Indeed, [Linder \(2001\)](#) has suggested that all the control strains available should be used as the wild type controls for a transgenic or knock-out mouse model. Furthermore, the background for many disease models may be dynamic in nature. Many transgenic and knock-out mice are generated on a mixed background, and with subsequent generations any number of alleles, all potentially interacting with the gene under investigation, could be fixed into the final genome ([Hintze & Shesely 2002](#)). Thus, rapid and high through-put processes for phenotyping mouse models are set to become increasingly important to address such issues.

Many of the approaches used to generate a metabolic profile are relatively cheap on a per sample basis. For example, excluding the significant investment required to purchase the analytical equipment, the cost of preparing a sample for NMR spectroscopy is less than £1, and for metabolic profiling by GC-MS and LC-MS this cost is probably only three to four times larger. This allows the analysis of a wide range of tissues, organs or biofluids during a phenotyping study of a given model of disease. This was illustrated by [Griffin *et al.* \(2004a,b\)](#) when investigating a mouse model of Machado-Joseph disease (spinocerebellar ataxia), where as well as identifying metabolic deficits in the cerebellum, metabolic abnormalities were also identified in the cortex, a region of the brain not thought previously to be affected by the disease. Furthermore, the cheap sample costs allow the profiling of large populations, as illustrated by [Lenz *et al.* \(2004\)](#) who investigated dietary differences between Swedish and British healthy individuals across a total population of 180 individuals.

The relative ease in which metabolic biomarkers can be examined in a range of species has allowed the simultaneous comparison of several different species exposed to a common physiological or environmental stimulus or lesion. For example, [Bundy *et al.* \(2005\)](#) have investigated the impact of metal contamination on a range of worm species including *Lumbricus rubellus*,

Lumbricus terrestris and *Eisenia andrei*. Metabolites identified by ^1H NMR spectroscopy and PCA demonstrated that *L. rubellus* responded to a greater extent when compared with *E. andrei*, demonstrating how metabolic phenotypes can be used to identify which species are most affected by a given ecotoxicology lesion. Similarly, Bundy *et al.* (2004) have used metabolomics to investigate different *Bacillus cereus* strains to separate the microbes according to ecotype. Unlike genomic data, metabolomic analysis using ^1H NMR spectroscopy successfully separated the strains into laboratory strains and clinical strains. This approach could be widely applicable to general microbiology in order to identify different functional and physiological ecotypes on bacteria.

(b) *Transcriptomics and metabolomics*

(i) *Fatty liver disease*

Supplementation of orotic acid to normal food intake is known to induce fatty liver disease in a number of mammalian species including the rat. While this accumulation is similar to that induced by alcohol, and it is known that orotic acid supplementation disrupts the production of various Apo proteins, the exact mechanism from orotic acid exposure to disrupted lipid transport is unknown. To address this Griffin *et al.* (2004b) applied a combined genomic, transcriptomic and metabolomic approach to the problem.

While classically the out bred Wistar rat strain has been used to investigate orotic acid induced fatty liver disease, the in bred Kyoto strain was found to be more predisposed both at the transcriptional and metabolic level following a combined analysis of the liver tissue. After exposure to orotic acid, as well as the reduction in circulating LDL and VLDL lipids following reduced production of ApoB and ApoC previously reported in the literature, NMR-based metabolomics also detected increases in the concentration of β -hydroxybutyrate, as occurs during type II diabetes, and a decrease in phosphatidylcholine. Supplementation of the diet with adenosine completely reversed this phenotype.

To compare the solution and solid state spectra from the liver with the transcriptional profiles measured using DNA microarrays the pattern recognition process 'prediction to latent structures through partial least squares' was used to cross correlate the metabolite changes (X matrix) to the transcriptional dataset (Y matrix) (figure 5). This combined analysis identified metabolic changes involving uridine metabolism and choline turnover, and transcriptional changes such as the expression of stress proteins. As well as identifying certain targets for drug intervention, such as stearyl-CoA desaturase, the approach was able to in part model Kyoto rats as genomic perturbations of the Wistar rat. Furthermore, the definition of a 'metabotype' by NMR spectroscopy allowed the analysis of the transcriptional data in terms of the dynamic responses induced by the drug intervention.

(ii) *Hepatotoxicity induced by acetaminophen*

A combined transcriptional and metabolomic approach has been used to investigate acetaminophen induced hepatotoxicity (Coen *et al.* 2004). In this

study, the transcriptional profile of liver tissue was supplemented by metabolic profiles of liver tissue (both extracts and intact tissue), blood plasma and urine obtained using ^1H NMR spectroscopy. These metabolic profiles identified decreases in glucose and glycogen and increases in lipids in liver tissue, alongside glucose, pyruvate, alanine and lactate increases in blood plasma. These metabolic changes were indicative of an increased glycolytic flux, and this was also confirmed at the transcriptional level demonstrating that the two approaches provided complementary information in the drug safety assessment process.

(c) *Transcriptomics, proteomics and metabolomics*

A number of studies are underway to examine complex disease processes through the three major 'omic' approaches for profiling tissue. Such poly-omic approaches are particularly appropriate to diseases where conventional approaches have drawn a blank. One such example involves a combined transcriptomic, proteomic and metabolomic analysis of brain tissue from schizophrenics, bipolar patients and control tissue (Prabakaran *et al.* 2004; figure 6). This study made use of a brain bank managed by the Stanley Foundation, which ensured access to control tissue from age-matched individuals. Just as important the control tissue had a similar post-mortem delay in sampling, ensuring that any biochemical changes detected were not artefacts of the tissue storage. All three tiers of biological intermediates indicated a metabolic disorder associated with schizophrenia. Intriguingly the changes detected were more apparent in white matter compared to grey matter. However, one criticism of the interpretation of the results was that proteomics and metabolomics were often only used in a confirmatory role to support the changes detected transcriptionally. In order for studies to benefit from such global approaches to profiling mRNA, proteins and metabolites researchers will need to treat the results from the process equally, rather than reflect prior biases to a particular technology.

This systems biology approach has also been applied to understanding atherosclerosis in the ApoE knock-out mouse (Clish *et al.* 2004). In this mouse model human ApoE, a protein involved in the transport of fats in blood plasma, is over expressed in the mouse, preventing the clearance of lipids from the blood. While animals at nine weeks are phenotypically normal, by 25 weeks there are atherosclerotic plaques. Using LC-MS to profile lipids they were able to demonstrate altered fatty acid metabolism at 9 weeks in the presymptomatic mice, highlighting an increase in lipid triglycerides and a decrease in lysophosphocholines in ApoE knock-out mice. To examine these perturbations further Clish and colleagues investigated the alterations in the liver at the transcriptomic, proteomic and metabolomic levels. LC-MS identified a significant increase in certain lysophosphocholines, diacylglycerides, phosphocholine and triglycerides. From a total of 21 000 measured components they were also able to identify a range of transcripts, proteins and metabolites that were changed in the ApoE knock-out mouse. Finally, these changes were

Figure 5. Two methods for cross correlating data from different 'omic' approaches. (a) shows a heat map plot of correlation coefficients from a partial least squares model of 20 transcripts (along the x -axis) and 200 metabolites (along the y -axis). Regions with an intense red colour are positively correlated. (b) Histograms of bootstraps for correlation coefficients between key metabolite regions and transcripts. The x -axis represents the correlation coefficients while the y -axis represents the number of times this correlation was returned during 10 000 iterations. Key: SCD, stearoylCoA desaturase 1; ApoC, apolipoprotein C III; MVLC, mitochondrial very long-chain acyl CoA thioesterase; GPAT, glycerol 3 phosphate acyltransferase; FAC, fatty acid CoA ligase 4; CH=CH unsaturated lipid resonance, CH₂CH₂CH₂ saturated lipid resonance (taken from Griffin *et al.* 2004b).

used to produce a correlation network, highlighting which metabolic pathways were most perturbed.

5. METABOLOMIC DATABASES

To a degree databases already exist that contain metabolic profiles or fingerprints in the form of large scale clinical chemistry assays or *in vivo* magnetic resonance spectra. This approach has been particularly useful at diagnosing and distinguishing different types of cancer. Pattern recognition has been used to simplify large sets of human MRS spectra for over a decade (Howells *et al.* 1992; Tate *et al.* 1996b; Usenius *et al.*

1996; Hagberg 1998; Preul *et al.* 1998; Gerstle *et al.* 2000). Furthermore, these approaches have also been used to identify metabolic fingerprints for a given tumour type (Gray *et al.* 1998; Gribbestad *et al.* 1999). These intelligent systems are now being used in cross centre studies such as that centred around the INTERPRET NMR database, developed as part of a collaboration between a number of European hospitals (Underwood *et al.* 2001; Howe *et al.* 2003). An automated pattern recognition approach has been implemented to help radiologists categorize magnetic resonance spectroscopy data of brain tumours

Figure 6. Metabolomic analysis of brain tissue from schizophrenics using high-resolution magic angle spinning ^1H NMR spectroscopy. Tissue was profiled using both a solvent suppressed pulse sequence (*a*, grey matter; *b*, white matter) and a T_2 relaxation weighted pulse sequence (A Carr Purcell Meiboom and Gill sequence) to selectively attenuate lipid resonances relative to aqueous metabolites (*c*, white matter). Spectra were acquired at 700.13 MHz at 3 °C and at a rotor spinning speed at 7000 Hz. This data was examined using multivariate data analysis including partial least squares-discriminate analysis (PLS-DA, a supervised regression extension of principal components analysis). The observation plot (*d*) of the PLS-DA model demonstrated that spectra of white matter from schizophrenic patients (open circle) formed a cluster apart from tissue from control patients (filled square) (cluster highlighted with broken line). The metabolic difference causing this was identified in the loadings plot (*e*) of the PLS-DA model, and was largely caused by increases in concentration of lactate (δ 1.32, 4.12), $-\text{CH}_2\text{CH}_2\text{CH}_2-$ lipid group (δ 1.36–1.32), phosphocholine (δ 3.24, 3.68) and *N*-acetyl aspartate (NAA) (δ 2.04), and decreases in CH_3- terminal lipid groups (δ 0.96–1.04) and myo-inositol (δ 3.52–3.60, 4.08), where δ signifies the centre of the 0.04 p.p.m. chemical shift region used as a variable in the multivariate analysis. A similar PLS-DA model could be built for solvent suppressed spectra from grey matter (*f*). Key: 1. $-\text{CH}_2\text{CH}_3$ lipid group, 2. leucine, isoleucine and valine, 3. lactate (sharp doublet) superimposed on $-\text{CH}_2\text{CH}_2\text{CH}_2-$ lipid resonance (broad resonance), 4. alanine, 5. acetate, 6. NAA, 7. glutamate and glutamine, 8. citrate, 9. creatine, 10. choline, 11. phosphocholine, 12. phosphatidylcholine and glycerophosphocholine, 13. taurine, 14. myo-inositol (series of resonances from 3.52 to 3.60). (*g*) A diagrammatic summary of the transcriptional changes identified alongside the metabolomic analysis. Numbers signify the number of transcripts identified as increased or decreased in expression in each group (taken from Prabakaran *et al.* 2004).

according to histological type and grade. It can successfully distinguish meningiomas, low-grade astrocytomas, and 'aggressive tumours' (glioblastomas and metastases). This pattern recognition process was even successful at predicting tumour types using data acquired at another site using a different magnetic resonance imaging instrument, indicating pattern recognition algorithms are less sensitive to acquisition parameters than had been expected.

The development of central databases for high-resolution/multivariate metabolomic data, that uses information from several different sites, will require the development of a mark up language and an ontology that completely describes a metabolomic experiment. Such languages have already been described for transcriptomics in the form of MIAME and proteomics in the form of MIAPE (Brazma *et al.* 2001; Taylor *et al.* 2003). Jenkins and colleagues have described a data model for plant metabolomics referred to as ArMet (architecture for metabolomics) (Jenkins *et al.* 2004; www.armet.org). This approach reflected in part the authors' biases towards GC-MS and plant metabolomics. In contrast, a standard metabolomic reporting structure (SMRS; www.smrs.org) has been proposed by those largely involved in toxicology/NMR based metabolomics. Of course, there is much common ground between these two approaches and it is hoped that a common formal description can be produced to allow the rapid comparison and exchange of metabolomic data in the future.

6. PREPARING FOR THE BALL: LOOKING TO THE FUTURE

(a) *New technology*

One of the key areas for improvement if metabolomics is to realize its primary objective of providing complete coverage of a cell, tissue, organ or organism metabolism is that metabolomic technology must be increased in terms of metabolite coverage. With metabolites in the cell spread across a concentration range of 10^9 , and a log polarity range of ~ -6 to 14 this is a huge challenge. To address this challenge a wide range of technological approaches are currently being examined. In addition to NMR and mass spectrometry based approaches researchers are using a wide range of other approaches including thin layer chromatography, coulometric arrays and metabolite arrays. For example, Kristal (2004) is using liquid chromatography in conjunction with coulometric arrays to identify discrete serotypes in mice undergoing dietary restriction compared with those animals fed a normal diet.

The popularity of LC-MS and GC-MS as metabolic profiling tools is set to increase as the chromatography becomes more reliable, and the software for matching mass fragmentation patterns is improved. Indeed, a number of manufacturers are developing the necessary software to allow confident retention time matching in LC-MS data, allowing the import of data directly into pattern recognition software. As well as GC-MS and LC-MS, a number of researchers are examining the advantages of also using other types of mass spectrometry including Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS;

Brown *et al.* 2005), matrix assisted laser desorption ionisation (MALDI)-MS and capillary electrophoresis-mass spectrometry (Sato *et al.* 2004).

There are also technologies being developed to make NMR spectroscopy a more sensitive approach both in terms of the sample volume examined and the number of metabolites that can be identified. Miniaturization has allowed the analysis of very small samples, including the CSF of a mouse ($\sim 1-3 \mu\text{l}$ of volume of fluid; Griffin *et al.* 2003). Similarly, Khandelwal *et al.* (2004) have investigated tetrodotoxin toxicity in the rat frontal cortex using microdialysis and microprobe NMR analysis. Tetrodotoxin induced decreases in glutamate, isoleucine, valine, alanine and α - and β -hydroxybutyrate suggested that these metabolites were normally released by neurons. In addition, a number of metabolic studies have been carried out on tissue dissected from laser capture arrangement. This demonstrates that NMR spectroscopy can examine very small sample sizes.

To date NMR-based techniques have focused on using ^1H NMR spectroscopy, but this approach suffers from a small chemical shift range, producing significant overlap of the resonances of a number of different metabolites. While ^{13}C NMR spectroscopy has a much larger chemical shift range, allowing the resolution of a wider range of metabolites, the approach is intrinsically less sensitive compared with ^1H NMR spectroscopy, as a result of the lower gyromagnetic ratio of the ^{13}C nucleus compared with the ^1H . However, in cryoprobes, a NMR probe where the receiver and transmitter coil is cooled using liquid helium, a significant improvement in sensitivity can be achieved by cooling the coil of a NMR probe to $\sim 4 \text{ K}$, allowing the rapid acquisition of ^{13}C NMR spectra. Keun *et al.* (2002) have already applied this approach to studying hydrazine toxicity through biofluid ^{13}C NMR spectroscopy. While in this particular study the biomarkers of hydrazine toxicity were already largely known, the use of ^{13}C spectroscopy did allow the identification of these metabolites largely from one-dimensional spectroscopy, without the need to identify the metabolites responsible for key resonances from a series of two-dimensional approaches. Furthermore, this approach may be particularly good for quantifying concentrations of metabolites which only produce singlets in ^1H NMR spectra. While ^1H NMR may not conclusively identify the metabolite, the extra chemical shift range for ^{13}C NMR is usually enough to allow unambiguous assignment of a given singlet.

One study that shows the potential of ^{13}C NMR spectroscopy was carried out by Boros *et al.* (2003), describing a technique for stable isotope-based dynamic metabolic profiling (SIDMAP). The approach relies on using $(1,2-^{13}\text{C}_2)\text{glucose}$ to label a proportion of the metabolome to identify which metabolic pathways are most active. Proliferating tumours have a distinct metabolic phenotype characterized by increased and preferential utilization of glucose through the non-oxidative pathway of the pentose cycle for nucleic acid synthesis, and a reduced rate of *de novo* fatty acid synthesis and TCA cycle glucose oxidation. By using steady state labellings of the pentose phosphate pathway intermediates as biomarkers of cell proliferation it is

hoped that such approaches could be extended to assessing treatments targeted at cellular proliferation, and in particular to understanding why some tumours are resistant to treatment by drugs such as Gleevac.

Another approach to improve the sensitivity of the NMR experiment is to hyphenate the NMR spectroscopy with liquid chromatography. This improves sensitivity by two mechanisms. Firstly, high- and low-concentration metabolites are separated by the liquid chromatography, reducing the likelihood of co-resonant peaks and also improving the dynamic range of the NMR experiment for the low-concentration metabolites. Secondly, metabolites are concentrated by the chromatography, further aiding the detection of low-concentration metabolites. Bailey *et al.* (2002, 2003) have used this approach of LC-NMR spectroscopy to metabolically profile a number of plants. This may be a particularly useful approach if hyphenated further with cryoprobe technology to allow cryo-LC-NMR.

The development of *in vivo* MRS continues, and with improved coil design and better localization pulse sequences spectra are obtainable that are comparable in resonance line width to those obtainable using ^1H MAS NMR spectroscopy, particularly for fatty tissue. Even for cerebral tissue a large number of metabolites can now be detected and quantified *in vivo*, with Pfeuffer *et al.* (1999) detecting 18 metabolites in the human brain. Thus, the ultimate future for the tool may be as an *in vivo* functional genomic tool.

Perhaps the biggest challenge will come in terms of the data integration that will be necessary in order to cross correlate information obtained from all these different approaches. However, only by combining technologies will we be able to expand our coverage of an organism's metabolome. Thus, there is an urgent need for improved pattern recognition processes for integrating the information produced by a variety of analytical approaches to provide a fuller coverage of the metabolome. Van der Greef (2004) is currently applying such global analytical approaches using a combination of NMR, GC-MS, LC-MS and FT-MS to profile the maximum number of metabolites in a tissue or biofluid. The techniques include both non-directed and class specific analysis of the medium, including specific analysis of fatty acids, amino acids, steroids, peptides and trace elements. The strategy of the process is to identify metabolic profiles associated with the disease before identifying unique biomarkers. Similarly, Gamache *et al.* (2004) have used a combination of LC-MS and HPLC electrochemical array (EC-array) to examine xenobiotic toxicity in rats as measured through urinary analysis. This approach allowed the quantification of metabolites at the fmol level. They also suggest the use of EC reactor cells in series with MS so as to facilitate structural elucidation.

(b) *New software for metabolomics*

Software development for metabolomics is occurring on two fronts. The development of improved pattern recognition and multivariate statistics approaches are needed to generate biomarkers and metabolite profiles associated with a process. There is also a need to develop improved peak alignment software for identifying metabolites in chromatograms, mass and NMR

spectra. One such algorithm increased the number of peaks detected in ^1H NMR spectra of urine from ~240 spectral regions of fixed widths to ~1000 individual peaks (Torgrip *et al.* 2003). Similarly, Cloarec *et al.* (2005) have developed an approach that takes into consideration the variability associated with pH and chelation of cations to improve classification in PCA models of NMR spectra from analysis of urine from drug toxicity studies.

There is also a need to integrate the data obtained from metabolomics with more classical approaches to metabolism. Metabolic pathways have been at the centre of our understanding of metabolism and there is a need to relate changes in discrete metabolite levels to global changes across metabolic pathways and networks, so that the causes of a perturbation can be separated from the collateral damage. Lange & Ghassemian (2005) have developed a visual interface, which combines biochemical pathway information with a gene-function database, which could be combined with transcriptional, proteomic and metabolomic data for *Arabidopsis thaliana*. Similarly, Romero *et al.* (2005) have used computational analysis of the human genome to identify possible unidentified enzymes in human metabolism. These 'pathway holes' were calculated to number 203, and 25 of these were assigned to putative genes.

7. CONCLUSIONS

While metabolomic approaches have figured to a lesser extent in functional genomics compared with transcriptomics and proteomics, some of the recent successes using global profiling tools for metabolism indicate that it will be increasingly used both in its own right and as a means for cross correlating information from genomics, transcriptomics and proteomics. Furthermore, metabolomics may define at which time points metabolism is most perturbed, highlighting where other functional genomic tools should focus. Mathematical tools will be needed for data fusion across the 'omic' platforms to achieve this. Metabolomics is also set to be used increasingly for phenotyping organisms, especially following the proliferation of techniques used to genetically modify an organism such as gene knock outs, gene knock ins and RNAi. This increased interest will necessitate metabolomic databases if information is to be passed between research groups, in turn necessitating a MIAME scheme for metabolomics. Currently, there is no consensus as to what information should be reported alongside a metabolomic study so that the data can be interpreted by other researchers. Such metadata (i.e. data about the data) will be vital if researchers are to generate databases akin to those being produced by the microarray and proteomic communities. If metabolomics is to attend the functional genomic ball it will be necessary to 'dress' in similar finery to that worn by other omic approaches.

REFERENCES

- Allen, J., Davey, H. M., Broadhurst, D., Heald, J. K., Rowland, J. J., Oliver, S. G. & Kell, D. B. 2003 High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat. Biotechnol.* **21**, 692–696. (doi:10.1038/nbt823)

- Bailey, N. J., Sampson, J., Hylands, P. J., Nicholson, J. K. & Holmes, E. 2002 Multi-component metabolic classification of commercial feverfew preparations via high-field ^1H -NMR spectroscopy and chemometrics. *Planta Med.* **68**, 734–738. (doi:10.1055/s-2002-33793)
- Bailey, N. J., Oven, M., Holmes, E., Nicholson, J. K. & Zenk, M. H. 2003 Metabolomic analysis of the consequences of cadmium exposure in *Silene cucubalus* cell cultures via ^1H NMR spectroscopy and chemometrics. *Phytochemistry* **62**, 851–858. (doi:10.1016/S0031-9422(02)00719-7)
- Beckwith-Hall, B. M., Nicholson, J. K., Nicholls, A. W., Foxall, P. J., Lindon, J. C., Connor, S. C., Abdi, M., Connelly, J. & Holmes, E. 1998 Nuclear magnetic resonance spectroscopic and principal components analysis investigations into biochemical effects of three model hepatotoxins. *Chem. Res. Toxicol.* **11**, 260–272. (doi:10.1021/tx9700679)
- Bendall, J. K., Heymes, C., Wright, T. J. F., Wheatcroft, S., Grieve, D. J., Shah, A. M. & Cave, A. C. 2002 *J. Mol. Cell. Cardiol.* **34**, 1325–1333. (doi:10.1006/jmcc.2002.2083)
- Berger A. 2004 Cambridge Healthtech Institute Conf. on Metabolic Profiling: Pathways in discovery. Florida, USA.
- Boros, L. G., Brackett, D. J. & Harrigan, G. G. 2003 Metabolic biomarker and kinase drug target discovery in cancer using stable isotope-based dynamic metabolic profiling (SIDMAP). *Curr. Cancer Drug Targets* **3**, 445–453. (doi:10.2174/1568009033481769)
- Brazma, A. *et al.* 2001 Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.* **29**, 365–371. (doi:10.1038/ng1201-365)
- Brindle, K. M. 2003 Metabolomics: Pandora's Box or Aladdin's Cave? *Biochemist* **25**, 15–17.
- Brindle, J. T. *et al.* 2002 Rapid and non-invasive diagnosis of the presence and severity of coronary heart disease using ^1H -NMR-based metabolomics. *Nat. Med.* **8**, 1439–1444. (doi:10.1038/nm802)
- 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. *Analyst* **128**, 32–36. (doi:10.1039/b209155k)
- Brown, S. C., Kruppa, G. & Dasseux, J. L. 2005 Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrom. Rev.* **24**, 223–231. (doi:10.1002/mas.20011)
- Bundy, J. G., Spurgeon, D. J., Svendsen, C., Hankard, P. K., Weeks, J. M., Osborn, D., Lindon, J. C. & Nicholson, J. K. 2004 Environmental metabolomics: applying combination biomarker analysis in earthworms at a metal contaminated site. *Ecotoxicology* **13**, 797–806. (doi:10.1007/s10646-003-4477-1)
- Bundy, J. G., Willey, T. L., Castell, R. S., Ellar, D. J. & Brindle, K. M. 2005 Discrimination of pathogenic clinical isolates and laboratory strains of *Bacillus cereus* by NMR-based metabolomic profiling. *FEMS Microbiol. Lett.* **242**, 127–136. (doi:10.1016/j.femsle.2004.10.048)
- Burton, E. A. & Davies, K. E. 2002 Muscular dystrophy: reason for optimism? *Cell* **108**, 5–8. (doi:10.1016/S0092-8674(01)00626-2)
- Chen, J.-H., Enloe, B. M., Fletcher, C. D., Cory, D. G. & Singer, S. 2001 Biochemical analysis using high-resolution magic angle spinning NMR spectroscopy distinguishes lipoma-like well-differentiated liposarcoma from normal fat. *J. Am. Chem. Soc.* **123**, 9200–9201. (doi:10.1021/ja016182u)
- Cheng, L. L., Lean, C. L., Bogdanova, A., Wright, S. C., Ackerman, J. L., Brady, T. J. & Garrido, L. 1996 Enhanced resolution of proton NMR spectra of malignant lymph nodes using magic angle spinning. *Magn. Reson. Med.* **36**, 653–658.
- Clish, C. B. *et al.* 2004 Integrative biological analysis of the APOE*3-leiden transgenic mouse. *Omic* **8**, 3–15. (doi:10.1089/153623104773547453)
- Cloarec, O., Dumas, M. E., Trygg, J., Craig, A., Barton, R. H., Lindon, J. C., Nicholson, J. K. & Holmes, E. 2005 Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in ^1H NMR spectroscopic metabolomic studies. *Anal. Chem.* **77**, 517–526. (doi:10.1021/ac048803i)
- Coen, M., Ruepp, S. U., Lindon, J. C., Nicholson, J. K., Pognan, F., Lenz, E. M. & Wilson, I. D. 2004 An integrated metabolomic investigation of acetaminophen toxicity in the mouse using NMR spectroscopy. *J. Pharm. Biomed. Anal.* **35**, 93–105. (doi:10.1016/j.jpba.2003.12.019)
- Connor, S. C., Hodson, M. P., Ringeissen, S., Sweatman, B. C., McGill, P. J., Waterfield, C. J. & Haselden, J. N. 2004 Development of a multivariate statistical model to predict peroxisome proliferation in the rat, based on urinary ^1H -NMR spectral patterns. *Biomarkers* **9**, 364–385. (doi:10.1080/13547500400006005)
- De Luca, V. & St Pierre, B. 2000 The cell and developmental biology of alkaloid biosynthesis. *Trends Plant Sci.* **5**, 168–173. (doi:10.1016/S1360-1385(00)01575-2)
- Emery, A. E. 1989 Emery–Dreifuss muscular dystrophy and other related disorders. *Br. Med. Bull.* **45**, 772–787.
- Fiehn, O. 2001 Combining genomics, metabolome analysis and biochemical modeling to understand metabolic networks. *Comp. Funct. Genom.* **2**, 155–168. (doi:10.1002/cfg.82)
- Fiehn, O. 2002 Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**, 155–171. (doi:10.1023/A:1013713905833)
- Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R. N. & Willmitzer, L. 2000 Metabolite profiling for plant functional genomics. *Nat. Biotech.* **18**, 1157–1161. (doi:10.1038/81137)
- Florian, C. L., Preece, N. E., Bhakoo, K. K., Williams, S. R. & Noble, M. D. 1995a Characteristic metabolic profiles revealed by ^1H NMR spectroscopy for three types of human brain and nervous system tumours. *NMR Biomed.* **8**, 253–264.
- Florian, C. L., Preece, N. E., Bhakoo, K. K., Williams, S. R. & Nobel, M. D. 1995b Cell type-specific fingerprinting of meningioma and meningeal cells by proton nuclear magnetic resonance spectroscopy. *Cancer Res.* **55**, 420–427.
- Gamache, P. H., Meyer, D. F., Granger, M. C. & Acworth, I. N. 2004 Metabolomic applications of electrochemistry/mass spectrometry. *J. Am. Soc. Mass Spectrom.* **15**, 1717–1726. (doi:10.1016/j.jasms.2004.08.016)
- Gavaghan, C. L., Holmes, E., Lenz, E., Wilson, I. D. & Nicholson, J. K. 2000 An NMR-based metabolomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse. *FEBS Lett.* **484**, 169–174. (doi:10.1016/S0014-5793(00)02147-5)
- Gerstle, R. J., Aylward, S. R., Kromhout-Schiro, S. & Mukherji, S. K. 2000 The role of neural networks in improving the accuracy of MR spectroscopy for the diagnosis of head and neck squamous cell carcinoma. *Am. J. Neuroradiol.* **21**, 1133–1138.
- Gray, H. F., Maxwell, R. J., Martinez-Perez, I., Arus, C. & Cerdan, S. 1998 Genetic programming for classification and

- feature selection: analysis of ^1H nuclear magnetic resonance spectra from human brain tumour biopsies. *NMR Biomed.* **11**, 217–224. (doi:10.1002/(SICI)1099-1492(199806/08)11:4/5<217::AID-NBM512>3.0.CO;2-4)
- Gribbestad, I. S., Sitter, B., Lundgren, S., Krane, J. & Axelson, D. 1999 Metabolite composition in breast tumors examined by proton nuclear magnetic resonance spectroscopy. *Anticancer Res.* **19**, 1737–1746.
- Griffin, J. L. & Bollard, M. E. 2004 Metabonomics: its potential as a tool in toxicology for safety assessment and data integration. *Curr. Drug Metab.* **5**, 389–398.
- Griffin, J. L., Williams, H. J., Sang, E., Clarke, K., Rae, C. & Nicholson, J. K. 2001 Metabolic profiling of genetic disorders: a multitissue (^1H) nuclear magnetic resonance spectroscopic and pattern recognition study into dystrophic tissue. *Anal. Biochem.* **293**, 16–21. (doi:10.1006/abio.2001.5096)
- Griffin, J. L., Sang, E., Evens, T., Davies, K. & Clarke, K. 2002 Metabolic profiles of dystrophin and utrophin expression in mouse models of Duchenne muscular dystrophy. *FEBS Lett.* **530**, 109–116. (doi:10.1016/S0014-5793(02)03437-3)
- Griffin, J. L., Lehtimäki, K. K., Valonen, P. K., Grohn, O. H., Kettunen, M. I., Ylä-Herttuala, S., Pitkanen, A., Nicholson, J. K. & Kauppinen, R. A. 2003 Assignment of ^1H nuclear magnetic resonance visible polyunsaturated fatty acids in BT4C gliomas undergoing ganciclovir-thymidine kinase gene therapy-induced programmed cell death. *Cancer Res.* **63**, 3195–3201.
- Griffin, J. L., Cemal, C. K. & Pook, M. A. 2004a Defining a metabolic phenotype in the brain of a transgenic mouse model of spinocerebellar ataxia 3. *Physiol. Genom.* **16**, 334–340. (doi:10.1152/physiolgenomics.00149.2003)
- Griffin, J. L., Bonney, S. A., Mann, C., Hebbachi, A. M., Gibbons, G. F., Nicholson, J. K., Sholders, C. C. & Scott, J. 2004b An integrated reverse functional genomic and metabolic approach to understanding orotic acid-induced fatty liver. *Physiol. Genom.* **17**, 140–149. (doi:10.1152/physiolgenomics.00158.2003)
- Griffiths, J. R. & Stubbs, M. 2003 Opportunities for studying cancer by metabolomics: preliminary observations on tumors deficient in hypoxia-inducible factor 1. *Adv. Enzyme Regul.* **43**, 67–76. (doi:10.1016/S0065-2571(02)00030-4)
- Griffiths, J. R. *et al.* 2002 Metabolic changes detected by *in vivo* magnetic resonance studies of HEPA-1 wild-type tumor deficient in hypoxia-inducible factor-1 β (HIF-1 β): evidence of an anabolic role for the HIF-1 pathway. *Cancer Res.* **62**, 688–695.
- Gullberg, J., Jonsson, P., Nordstrom, A., Sjöström, M. & Moritz, T. 2004 Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry. *Anal. Biochem.* **331**, 283–295. (doi:10.1016/j.ab.2004.04.037)
- Hagberg, G. 1998 From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods. *NMR Biomed.* **11**, 148–156. (doi:10.1002/(SICI)1099-1492(199806/08)11:4/5<148::AID-NBM511>3.0.CO;2-4)
- Hintze, T. H. & Shesely, E. G. 2002 Is a mouse like any other mouse? *J. Mol. Cell. Cardiol.* **34**, 1283–1286. (doi:10.1006/jmcc.2002.2103)
- Hoffman, E. P., Monaco, A. P., Feener, C. C. & Kunkel, L. M. 1987 Conservation of the Duchenne muscular dystrophy gene in mice and humans. *Science* **238**, 347–350.
- Holmes, E. *et al.* 1998 Development of a model for classification of toxin-induced lesions using ^1H NMR spectroscopy of urine combined with pattern recognition. *NMR Biomed.* **11**, 235–244. (doi:10.1002/(SICI)1099-1492(199806/08)11:4/5<235::AID-NBM507>3.0.CO;2-V)
- Howe, F. A. *et al.* 2003 Metabolic profiles of human brain tumors using quantitative *in vivo* ^1H magnetic resonance spectroscopy. *Magn. Reson. Med.* **49**, 223–232. (doi:10.1002/mrm.10367)
- Howells, S. L., Maxwell, R. J., Peet, A. C. & Griffiths, J. R. 1992 An investigation of tumor ^1H nuclear magnetic resonance spectra by the application of chemometric techniques. *Magn. Reson. Med.* **28**, 214–236.
- Huxtable, R. J. 1992 Physiological actions of taurine. *Physiol. Rev.* **72**, 101–163.
- Jenkins, H. *et al.* 2004 A proposed framework for the description of plant metabolomics experiments and their results. *Nat. Biotech.* **22**, 1601–1606. (doi:10.1038/nbt1041)
- Jones, G. L. *et al.* 2005 A functional analysis of mouse models of cardiac disease through metabolic profiling. *J. Biol. Chem.* **280**, 7530–7539. (doi:10.1074/jbc.M410200200)
- 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 metabolomic studies. *Anal. Chem.* **74**, 4588–4593. (doi:10.1021/ac025691r)
- Khandelwal, P., Beyer, C. E., Lin, Q., Schechter, L. E. & Bach 2nd, A. C. 2004 Studying rat brain neurochemistry using nanoprobe NMR spectroscopy: a metabolomics approach. *Anal. Chem.* **76**, 4123–4127. (doi:10.1021/ac049812u)
- Kristal, B. 2004 Cambridge Healthtech Institute Conf. on Metabolic Profiling: Pathways in discovery. Florida, USA, Dec 2004.
- Lange, B. M. & Ghassemian, M. 2005 Comprehensive post-genomic data analysis approaches integrating biochemical pathway maps. *Phytochemistry* **66**, 413–451. (doi:10.1016/j.phytochem.2004.12.020)
- Lehtimäki, K. K. *et al.* 2003 Metabolite changes in BT4C rat gliomas undergoing ganciclovir-thymidine kinase gene therapy-induced programmed cell death as studied by ^1H NMR spectroscopy *in vivo*, *ex vivo*, and *in vitro*. *J. Biol. Chem.* **278**, 45 915–45 923. (doi:10.1074/jbc.M306209200)
- Lenz, E. M., Bright, J., Wilson, I. D., Hughes, A., Morrisson, J., Lindberg, H. & Lockton, A. 2004 Metabonomics, dietary influences and cultural differences: a ^1H NMR-based study of urine samples obtained from healthy British and Swedish subjects. *Pharm. Biomed. Anal.* **36**, 841–849. (doi:10.1016/j.jpba.2004.08.002)
- Linder, C. C. 2001 The influence of genetic background on spontaneous and genetically engineered mouse models of complex diseases. *Lab. Anim.* **30**, 34–39.
- McIntosh, L., Granberg, K. E., Briere, K. M. & Anderson, J. E. 1998a Nuclear magnetic resonance spectroscopy study of muscle growth, mdx dystrophy and glucocorticoid treatments: correlation with repair. *NMR Biomed.* **11**, 1–10. (doi:10.1002/(SICI)1099-1492(199802)11:1<1::AID-NBM493>3.3.CO;2-4)
- McIntosh, L. M., Garrett, K. L., Megeney, L., Rudnicki, M. A. & Anderson, J. E. 1998b Regeneration and myogenic cell proliferation correlate with taurine levels in dystrophin- and MyoD-deficient muscles. *Anat. Rec.* **252**, 311–324. (doi:10.1002/(SICI)1097-0185(199810)252:2<311::AID-AR17>3.0.CO;2-Q)
- Millis, K., Weybright, P., Cambell, N., Fletcher, J. A., Fletcher, C. D., Cory, D. G. & Singer, S. 1999 Classification of human liposarcoma and lipoma using *ex*

- in vivo* proton NMR spectroscopy. *Magn. Reson. Med.* **41**, 257–267. (doi:10.1002/(SICI)1522-2594(199902)41:2<257::AID-MRM8>3.0.CO;2-N)
- 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**, 165–173. (doi:10.1021/tx034123j)
- Nicholson, J. K., Higham, D. P., Timbrell, J. A. & Sadler, P. J. 1989 Quantitative high resolution ¹H NMR urinalysis studies on the biochemical effects of cadmium in the rat. *Mol. Pharmacol.* **36**, 398–404.
- Nicholson, J. K., Lindon, J. C. & Holmes, E. 1999 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29**, 1181–1189. (doi:10.1080/004982599238047)
- Nicholson, J. K., Connelly, J., Lindon, J. C. & Holmes, E. 2002 Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* **1**, 153–161. (doi:10.1038/nrd728)
- Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. 1998 Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **16**, 373–378. (doi:10.1016/S0167-7799(98)01214-1)
- Pfeuffer, J., Tkac, I., Provencher, S. W. & Gruetter, R. 1999 Towards an *in vivo* neurochemical profile: quantification of 18 metabolites in short-echo-time ¹H NMR spectra of the rat brain. *J. Magn. Reson.* **141**, 104–120. (doi:10.1006/jmre.1999.1895)
- Prabakaran, S. *et al.* 2004 Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress. *Mol. Psychiatry* **9**, 684–697, 643.
- Preul, M. C., Caramanos, Z., Leblanc, R., Villemure, J. G. & Arnold, D. L. 1998 Using pattern analysis of *in vivo* proton MRSI data to improve the diagnosis and surgical management of patients with brain tumors. *NMR Biomed.* **11**, 192–200. (doi:10.1002/(SICI)1099-1492(199806/08)11:4/5<192::AID-NBM535>3.0.CO;2-3)
- Raamsdonk, L. M. *et al.* 2001 A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* **19**, 45–50. (doi:10.1038/83496)
- Ringeissen, S. *et al.* 2004 Potential urinary and plasma biomarkers of peroxisome proliferation in the rat. *Biomarkers* **8**, 240–271. (doi:10.1080/1354750031000149124)
- Romero, P., Wagg, J., Green, M. L., Kaiser, D., Krummenacker, M. & Karp, P. D. 2005 Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.* **6**, R2. (doi:10.1186/gb-2004-6-1-r2) [Published online 22 December 2004.]
- 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**, 151–163. (doi:10.1111/j.1365-313X.2004.02187.x)
- Steuer, R., Kurths, J., Fiehn, O. & Weckwerth, W. 2003 Observing and interpreting correlations in metabolomic networks. *Bioinformatics* **19**, 1019–1026. (doi:10.1093/bioinformatics/btg120)
- Tanabe, Y., Esaki, K. & Nomura, T. 1986 Skeletal muscle pathology in X chromosome-linked muscular dystrophy (mdx) mouse. *Acta Neuropathol. (Berl)* **69**, 91–95. (doi:10.1007/BF00687043)
- Tate, A. R., Crabb, S., Griffiths, J. R., Howells, S. L., Mazucco, R. A., Rodrigues, L. M. & Watson, D. 1996a Lipid metabolite peaks in pattern recognition analysis of tumour *in vivo* MR spectra. *Anticancer Res.* **16**, 1575–1579.
- Tate, A. R., Watson, D., Eglen, S., Arvanitis, T. N., Thomas, E. L. & Bell, J. D. 1996b Automated feature extraction for the classification of human *in vivo* ¹³C NMR spectra using statistical pattern recognition and wavelets. *Magn. Reson. Med.* **35**, 834–840.
- Tate, A. R. *et al.* 1998 Towards a method for automated classification of ¹H MRS spectra from brain tumours. *NMR Biomed.* **11**, 177–191. (doi:10.1002/(SICI)1099-1492(199806/08)11:4/5<177::AID-NBM534>3.0.CO;2-U)
- Tate, A. R., Majos, C., Moreno, A., Howe, F. A., Griffiths, J. R. & Arus, C. 2003 Automated classification of short echo time *in vivo* ¹H brain tumor spectra: a multicenter study. *Magn. Reson. Med.* **49**, 29–36. (doi:10.1002/mrm.10315)
- Taylor, C. F. *et al.* 2003 A systematic approach to modeling, capturing, and disseminating proteomics experimental data. *Nat. Biotech.* **21**, 247–254. (doi:10.1038/nbt0303-247)
- Tinsley, J. M., Potter, A. C., Phelps, S. R., Fisher, R., Trickett, J. I. & Davies, K. E. 1996 Amelioration of the dystrophic phenotype of mdx mice using a truncated utrophin transgene. *Nature* **384**, 349–353. (doi:10.1038/384349a0)
- Tinsley, J., Deconinck, N., Fisher, R., Kahn, D., Phelps, S., Gillis, J. M. & Davies, K. 1998 Expression of full-length utrophin prevents muscular dystrophy in mdx mice. *Nat. Med.* **4**, 1441–1444. (doi:10.1038/4033)
- Tomlins, A., Foxall, P. J. D., Lindon, J. C., Lynch, M. J., Spraul, M., Everett, J. & Nicholson, J. K. 1998 High resolution magic angle spinning ¹H nuclear magnetic resonance analysis of intact prostatic hyperplastic and tumour tissues. *Anal. Commun.* **35**, 113–115. (doi:10.1039/a708098k)
- Torgrip, R. J. O., Aberg, M., Karlberg, B. & Jacobson, S. P. 2003 Peak alignment using reduced set mapping. *J. Chemometr.* **17**, 573–582. (doi:10.1002/cem.824)
- Underwood, J., Tate, A. R., Luckin, R., Majos, C., Capdevila, A., Howe, F., Griffiths, J. & Arus, C. 2001 A prototype decision support system for MR spectroscopy-assisted diagnosis of brain tumours. *Medinfo* **10**, 561–565.
- Usenius, J. P., Tuohimetsa, S., Vainio, P., Ala-Korpela, M., Hiltunen, Y. & Kauppinen, R. A. 1996 Automated classification of human brain tumours by neural network analysis using *in vivo* ¹H magnetic resonance spectroscopic metabolite phenotypes. *Neuroreport* **7**, 1597–1600.
- Valonen, P. K., Griffin, J. L., Lehtimäki, K. K., Liimatainen, T., Gröhn, O. H. J. & Kauppinen, R. A. 2005 High resolution magic angle spinning ¹H NMR spectroscopy of choline chemical shift region metabolites in a rat brain glioma undergoing gene therapy-induced programmed cell death. *NMR Biomed.* **18**, 252–259.
- van der Greef, J. 2004 Cambridge Healthtech Institute Conference on Metabolic Profiling: Pathways in discovery. Florida, USA.