

## Metabonomics in Diabetes Research

Johan H. Faber, Ph.D.,<sup>1</sup> Daniel Malmodin, Ph.D.,<sup>1</sup> Henrik Toft, Ph.D.,<sup>1</sup> Anthony D. Maher, Ph.D.,<sup>2</sup>  
Derek Crockford, Ph.D.,<sup>2</sup> Elaine Holmes, Ph.D.,<sup>2</sup> Jeremy K. Nicholson, Ph.D.,<sup>2</sup>  
Marc E. Dumas, Ph.D.,<sup>2</sup> and Dorrit Baunsgaard, Ph.D.<sup>1</sup>

### Abstract

Metabonomics has been defined as “quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification” and can provide information on disease processes, drug toxicity, and gene function. In this approach many samples of biological origin (biofluids such as urine or plasma) are analyzed using techniques that produce simultaneous detection. A variety of analytical metabolic profiling tools are used routinely, are also currently under development, and include proton nuclear magnetic resonance spectroscopy and mass spectrometry with a prior online separation step such as high-performance liquid chromatography, ultra-performance liquid chromatography, or gas chromatography. Data generated by these analytical techniques are often combined with multivariate data analysis, i.e., pattern recognition, for respectively generating and interpreting the metabolic profiles of the investigated samples. Metabonomics has gained great prominence in diabetes research within the last few years and has already been applied to understand the metabolism in a range of animal models and, more recently, attempts have been done to process complex metabolic data sets from clinical studies. A future hope for the metabonomic approach is the identification of biomarkers that are able to highlight individuals likely to suffer from diabetes and enable early diagnosis of the disease or the identification of those at risk. This review summarizes the technologies currently being used in metabonomics, as well as the studies reported related to diabetes prior to a description of the general objective of the research plan of the metabonomics part of the European Union project, Molecular Phenotyping to Accelerate Genomic Epidemiology.

*J Diabetes Sci Technol 2007;1(4):549-557*

**Author Affiliations:** <sup>1</sup>Novo Nordisk A/S, Novo Nordisk Park, Måløv, Denmark, and <sup>2</sup>Department of Biomolecular Medicine, Division of Surgery, Oncology, Reproductive Biology and Anaesthetics, Faculty of Medicine, Imperial College London, South Kensington, London, United Kingdom

**Abbreviations:** (ESI) electrospray ionization, (GC) gas chromatography, (HPLC) high-performance liquid chromatography, (MolPAGE) Molecular Phenotyping to Accelerate Genomic Epidemiology, (MS) mass spectrometry, (NAFLD) nonalcoholic fatty liver disease, (NMR) nuclear magnetic resonance, (PLS) partial least squares, (UPLC) ultraperformance liquid chromatography

**Keywords:** analytical tools, biomarkers, diabetes, metabonomics, multivariate analysis

**Corresponding Author:** Dorrit Baunsgaard, Ph.D., Novo Nordisk A/S, Novo Nordisk Park DK-2760 Måløv, Denmark; email address [doaba@novonordisk.com](mailto:doaba@novonordisk.com)

## Introduction

The identification of biomarkers is of growing interest in the fields of drug discovery, biological monitoring of toxic agents, and the study of disease disorders. Several technologies are being developed to improve the identification of such biomarkers, namely genomics and transcriptomics, which examine genetic complement and gene expression, respectively; proteomics, which involves the analysis of protein synthesis and cell signaling; and metabonomics, defined as the “quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification.”<sup>1</sup>

The term *-omics* represents the rigorous study of various collections of molecules, biological processes, or physiological functions and structures as systems. Whereas metabonomics provides a means for understanding the variation in low molecular mass metabolites in complex multicellular organisms and their response to change, the additional mentioned “*-omics*” sciences are concerned with cellular macromolecules.

The evolution of metabonomic technology in *in vivo* systems has to some extent been paralleled by the development of metabolomics in the field of cell biology. Metabolomics may be defined as the measurement of all the metabolites in a specified biological sample.<sup>2</sup>

In the approach of metabonomics, many samples of a biological origin are analyzed using techniques that produce simultaneous detection, thereby providing metabolite profiles. Metabonomics is mainly conducted on biofluids such as urine or plasma, which are easily obtainable in mammalian studies, but other fluids such as tissue extracts, *in vitro* cultures, and supernatants can also be used. Because biofluids fulfill diverse biological purposes, their metabolic composition varies with their role and the functional integrity of the organ systems that are communicating with them, and ultimately with the physiological status of the whole organism.<sup>3</sup>

The idea that complex metabolite profiles could be used to diagnose diseases in medicine was put forward early on by Linus Pauling and co-workers,<sup>3</sup> who suggested the analysis of vapor and breath condensates using gas chromatography (GC) in order to detect metabolic changes in indicative diseases. To investigate the complex metabolic consequences of disease processes, toxic reactions, and genetic manipulations, nonselective,

but specific, “information-rich” analytical approaches are required.

One of the most established methods for metabonomic research in mammalian studies to date is proton nuclear magnetic resonance (<sup>1</sup>H NMR) spectroscopy. This concept has arisen from work on the application of <sup>1</sup>H NMR spectroscopy to study the multicomponent metabolic composition of biofluids, cells, and tissues in the late 1970s and the early 1980s.<sup>4–7</sup> <sup>1</sup>H NMR studies of biofluids allow the simultaneous assessment of thousands of partially overlapping resonances arising from the presence of hundreds of endogenous molecules. The technique provides a characteristic <sup>1</sup>H NMR spectroscopic fingerprint of each biofluid sample in which the spectral intensity is determined by the relative concentrations of drug components present in the sample and, in some cases, by their intermolecular interactions.<sup>8,9</sup> <sup>1</sup>H NMR spectroscopy is very efficient in biochemical investigations, in that the analysis is nondestructive, nonselective, cost effective, and relatively sensitive (low-nanogram per milliliter detection limits are possible with the appropriate instrumentation).<sup>10</sup> Simple one-dimensional spectra typically take only a few minutes to acquire with no sample preparation other than buffering and the addition of heavy water (D<sub>2</sub>O) to provide a reference frequency.<sup>1</sup>

Advances in mass spectrometric (MS) techniques have resulted in the development of robust methods for the detection of molecular masses of organic molecules in biological matrices suitable for metabonomic studies.<sup>11</sup> GC hyphenated to MS has proved to be a valuable tool in metabolite profiling studies of plant metabolism.<sup>2,12</sup> However, there are limitations such as the size and type of metabolite that can be analyzed, and the metabolites in general have to be chemically derivatized, which makes the technique less suitable for metabonomic studies where a genuine high throughput is desired. The application of either high-performance liquid chromatography (HPLC) or ultraperformance liquid chromatography (UPLC) online with MS using electrospray ionization (ESI) within metabonomic research has become increasingly popular, particularly when combined with exact mass measurements, which allow elemental compositions to be determined. In contrast to GC-MS, the technique detects nonvolatile compounds, which make up a large proportion of metabolites. Furthermore, the technique offers potential advantages in terms of sensitivity

compared to NMR with sensitivities in picograms per milliliter readily achieved.<sup>13</sup> Unlike NMR, LC-ESI-MS resolves individual components into separate peaks, whereas NMR provides a chemical fingerprint, and MS metabonomics may also provide a means of detecting a different and complementary range of biomarkers.

In studies where both LC-MS and <sup>1</sup>H NMR have been used, it is clear that the combination offers advantages, especially for structural moieties that are invisible by NMR, e.g., sulfates, or not easily detected by MS, e.g., glucose.<sup>14,15</sup>

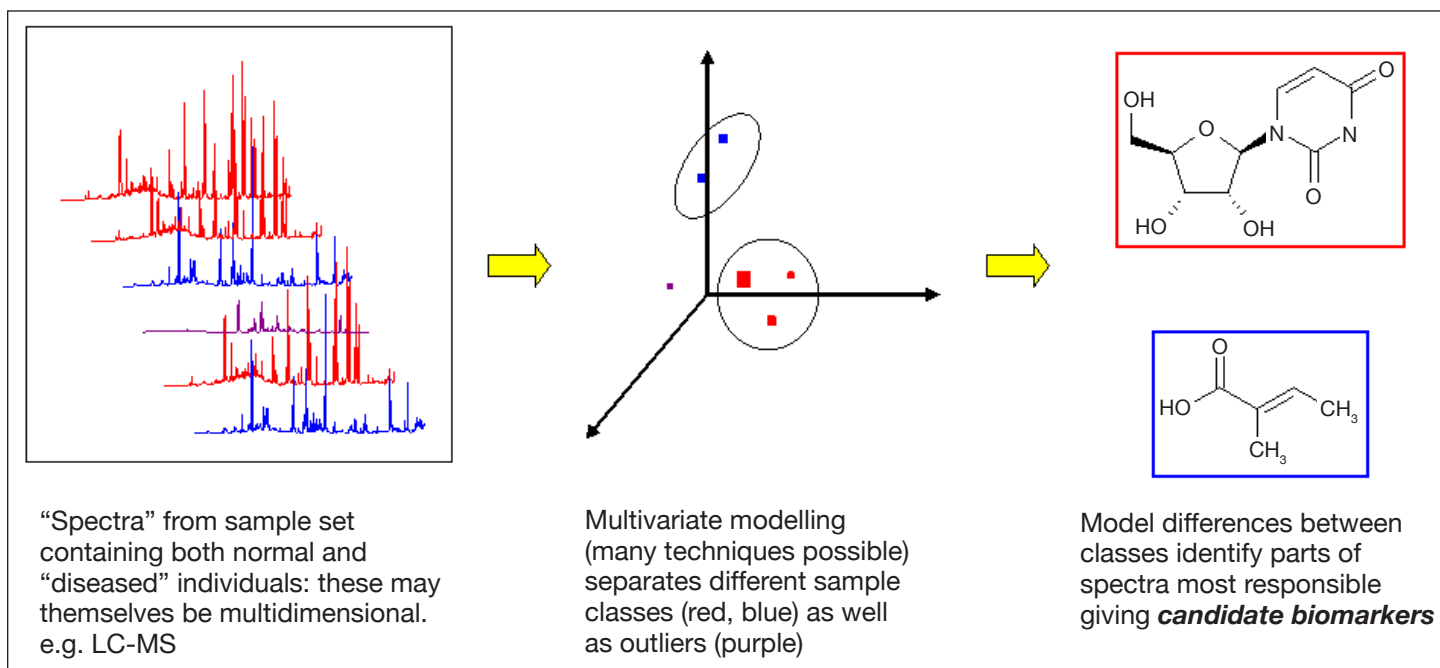
An alternative approach is the use of optical spectroscopic analysis, e.g., Fourier transform infrared and Raman spectroscopy, which are promising but still rather unexplored spectroscopic tools within metabonomic research.<sup>16,17</sup>

Although early metabolic profiling studies were conducted using instruments, which provided relatively low resolutions by today's standard, it was evident that the obtained metabolite profiles were rather complex.<sup>18,19</sup> As the different spectrometer technologies advanced, it became apparent that visual analysis was not the optimal way to interpret biofluid spectra, as the high level of metabolic information convolved in biofluid spectra presented a significant analysis challenge. Therefore, computer-aided pattern recognition and data reduction techniques were developed to classify biochemical changes that reflect specific physiological or pathological

states (see **Figure 1**).<sup>18,20</sup> In modern metabonomics research studies these techniques are often considered as integral parts and form a vital step in the identification of latent biomarkers.

Since the first applications of pattern recognition techniques on metabolite profiling data,<sup>18</sup> numerous approaches have been developed for the analysis of large spectral data sets, including principal component analysis-based metabolic trajectories,<sup>21</sup> partial least-squares (PLS)-based batch statistical process control,<sup>22</sup> hierarchical cluster analysis and *k*-nearest neighbor approaches,<sup>23</sup> a density superposition approach, "CLOUDS,"<sup>24</sup> statistical total correlation spectroscopy,<sup>25</sup> statistical heterospectroscopy,<sup>26</sup> multiple factor analysis,<sup>27</sup> and Bayesian spectral decomposition.<sup>28</sup> The major goal of these techniques is the classification of biofluid samples on the basis of differences between unaltered and "aberrant" spectral patterns related to the administration of a particular drug, toxin, or disease disorder, thus providing means of identifying biomarkers (see **Figure 1**).

Metabonomics studies of biofluids such as urine and plasma obtained from clinical trials have been applied successfully to investigate numerous diseases and metabolic processes.<sup>18,29–35</sup> Furthermore, the approach of metabonomics has proven sensitive enough for the understanding of gene function in model organisms, including yeast,<sup>36,37</sup> plants,<sup>2</sup> and mice.<sup>38</sup> This has led to the development of a number of screening assays in



**Figure 1.** Principles of a metabonomic study.

recent years to be used in drug development related to human diseases such as metabolic syndrome,<sup>39</sup> diabetes,<sup>40</sup> and cardiovascular disease,<sup>38</sup> as well as its potential use in toxicology as part of the drug safety assessment process.<sup>41</sup>

## Metabonomics in Diabetes Research

The total number of individuals suffering from type II diabetes worldwide has been projected to rise from 171 million in 2000 to 366 million in 2030.<sup>42</sup> Because of a lack of symptoms early on in the disease, a large proportion of these individuals remain undiagnosed; this proportion is estimated to be higher than 50%.<sup>43</sup> Indeed, diabetes care already accounts for about 2–7% of the total national health care budgets of western European countries.<sup>44</sup> Thus, identification of early biomarkers for prediction and monitoring is needed for adequate screening diagnostics of type II diabetes.

Several animal models have been developed and used extensively in studies of the pathophysiology of diabetes and its complications.<sup>45</sup> The obese Zucker (fa/fa) rat represents such an important model of type II diabetes.<sup>46,47</sup> Preliminary metabonomics investigations have examined different rat strains and obese Zucker (fa/fa) rats in order to obtain a better understanding of normal variation in biofluid composition and determine the differences between “normal” animals and obese Zucker (fa/fa) rats.<sup>40,48–50</sup> The analytical techniques used in these studies were able to detect discrimination among age, strain, gender, and diurnal variation of the animal models, which may provide biomarkers of diabetes.

The metabolite profiles of liver and blood from lean and obese Zucker (fa/fa) rats have been established and compared based on quantitative <sup>1</sup>H and phosphorus (<sup>31</sup>P) NMR analysis.<sup>51</sup> The study revealed metabolic abnormalities in the mitochondrial function and methionine metabolism for the obese Zucker (fa/fa) rats, which were found to result in a decreased hepatic energy state. <sup>1</sup>H NMR on liver extracts indicated significantly increased concentrations of fatty acids and triglycerides. From <sup>31</sup>P NMR spectra, a profound decrease in the adenosine 5'-triphosphate (ATP)/adenosine diphosphate ratio in the liver of obese Zucker (fa/fa) rats was demonstrated. Furthermore, a decrease in glutathione and a reduced ratio of polyunsaturated fatty acids: monosaturated fatty acids were found for the obese Zucker (fa/fa) rats.

Another study analyzed the plasma obtained from normal Wistar-derived and obese Zucker (fa/fa) rats

using <sup>1</sup>H NMR, GC-MS, and UPLC-MS analytical methodologies as part of metabonomics investigations of animal disease models.<sup>52</sup> All three techniques were found to provide the basis for methods to separate the plasma profiles of the two strains of rat. Some similarities were found between the classes of metabolites detected by GC-MS and <sup>1</sup>H NMR spectroscopy, whereas this was less apparent for metabolites detected using UPLC-MS. It was found that the combination of the three methods provided a comprehensive tool for metabolite profiling. Both GC-MS and <sup>1</sup>H NMR spectroscopy highlighted that glucose was present in lower concentrations in the plasma in obese Zucker (fa/fa) animals, and further on the role of elevations in the concentrations of lipids and sterols in these rats as biomarkers. The most significant markers identified by GC-MS were urea, arachidonic acid, and tocopherol. A large number of markers were detected by UPLC-MS, taurocholate was identified as being raised in Zucker animals, and the remaining markers detected were unidentified.

In a subsequent study by the same research groups using UPLC-MS to investigate the metabolic process in plasma from three different rat strains, the obese Zucker (fa/fa), Zucker lean, and the lean/(fa) obese cross, several markers were detected.<sup>53</sup> Again taurocholate was found to be raised in obese Zucker (fa/fa) rats. The additional markers detected in this study remained unidentified; however, they allowed a metabolic class separation of the three rat strains investigated.

Taurine is synthesized from the essential amino acid methionine; the three metabonomic studies mentioned earlier<sup>51–53</sup> are thus in agreement and suggest a change in taurine metabolism in the development of diabetes.

Highly complex animals such as mammals can be considered as “superorganisms” with a karyome, chondriome, and a microbiome, resulting from a genetic and metabolic coevolution process between symbiotic bacteria and the host.<sup>54</sup> Novel approaches are emerging to model metabolism in interacting cellular systems that also involve symbiotic microorganisms.<sup>55,56</sup> The intricate relationship has been investigated between gut microbiota and host cometabolic phenotypes associated with dietary-induced impaired glucose homeostasis and nonalcoholic fatty liver disease (NAFLD) in mouse strain 129S6,<sup>57</sup> known to be susceptible to these disease traits associated with type II diabetes.<sup>58</sup> Analysis of plasma and urine <sup>1</sup>H NMR-based metabonomic data showed that disruptions of choline metabolism related to gut microorganisms were associated with impaired glucose and NAFLD in 129S6. It was found that

conversion of choline into methylamines by microbiota in strain 129S6 on a high-fat diet reduces the bioavailability of choline and mimics the effect of choline-deficient diets causing NAFLD. This study highlights the metabolic role played by gut microbiota in the development of insulin resistance and thereby also type II diabetes, as observed for obesity recently.<sup>59,60</sup>

Some of the most effective drug therapies for type II diabetes treatment are based on the activation of peroxisome proliferator-activated receptors (PPAR).<sup>61</sup> The actions of these receptors have been investigated in metabonomic studies of mouse models,<sup>39,62</sup> and several methods based on the principles of metabonomics have been developed to easily evaluate the potential of clinical assessments of therapy of PPAR active drug candidates.<sup>63,64</sup>

Only a few metabonomics studies have been reported in the field of clinical trial research of diabetes mellitus.

The first initial study of the ability of <sup>1</sup>H NMR to provide useful "fingerprints" of the biochemical changes that accompany diseased states was obtained by recording <sup>1</sup>H NMR spectra of serum, plasma, and urine from fasting and type II diabetic subjects.<sup>33</sup> The levels of 3-D-hydroxybutyrate, acetoacetate, lactate, valine, and alanine were determined and compared between the two patient groups. In another study, <sup>1</sup>H NMR spectroscopy was used to examine serial urine samples from type II diabetic patients and control subjects.<sup>35</sup> Several metabolites were detected, and it was found that creatine, acetate, betaine, and ketone bodies were present more frequently and in greater concentrations in type II diabetic subjects than in controls. A similar approach was used in yet another study, where amniotic fluid obtained from third trimester pregnant women suffering from insulin-dependent diabetes mellitus and from control subjects was analyzed by <sup>1</sup>H NMR spectroscopy.<sup>34</sup> A subset of metabolites was identified, and for lactate, alanine, acetate, and glucose, a general diminution in the concentration was found, especially glucose in the diabetic group.

Mass spectrometric techniques have likewise been applied to obtain metabolite profiles of type II diabetic patients. In a preliminary study the metabonomics approach was applied in the diagnosis of type II diabetes.<sup>65</sup> It was found that a method based on serum lipid metabolite profiles obtained by GC-MS combined with pattern recognition analysis of data provided an effective approach to the discrimination of type II diabetic patients from healthy

controls. Subsequently, the same research group carried out a more detailed metabonomics study by applying HPLC-MS followed by PLS-discriminant analysis to study phospholipids in plasma of type II diabetic patients and controls.<sup>66</sup> The method allowed differentiation of the two groups investigated, and the type II diabetic patients were found to have an elevation of two phosphoethanolamine species and a decrease in two lysophosphocholine species.

## The Molecular Phenotyping to Accelerate Genomic Epidemiology (MolPAGE) Project

The European Union-funded consortium MolPAGE<sup>67</sup> was formed in 2004 involving 17 partners from universities, pharmaceutical companies, and biotechnology companies. The 4-year program aims to tackle diabetes and one of its major complications, vascular disease, at the level of genes, proteins, metabolites, and other biomarkers. The project brings together a range of techniques to characterize diabetes, which include metabonomics, transcriptomics, and proteomics.

The ultimate goal of the MolPAGE project is to identify biomarkers that are able to highlight individuals likely to suffer from diabetes and vascular disease in the future before they show any of the symptoms, biochemical abnormalities, or other features typically used in the diagnosis of these conditions. An early diagnosis of the disease or the identification of those at risk has the potential of allowing more effective prevention programs and better treatment of the disease.

Developments in metabonomics now enable much more detailed studies of metabolites involved in disease progression. The major opportunity for the project MolPAGE is to further develop the technology for high-throughput use to define the relationship between different biomarkers (and their patterns) and the disease on an epidemiological scale.

The program has three parts:

1. Evaluation of sample collection and storage methodology to reduce sample variation and improve stability.
2. Development of new tools for molecular phenotyping at the population level, including metabonomics, transcriptomics, and proteomics.
3. Development of new bioinformatics tools for data collection and integrated data analysis.

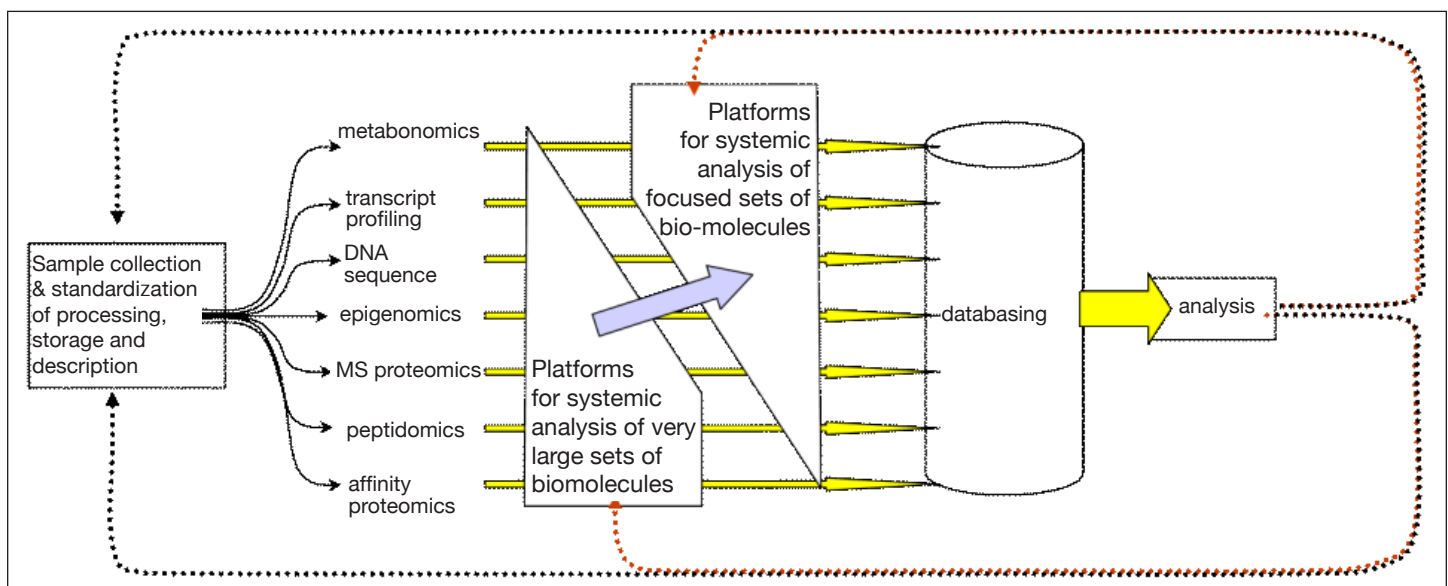
The access to appropriate clinical samples for testing, validation, and initial experiments is essential for the program to succeed. For the prosecution of metabonomics investigations within the MolPAGE project, *de novo* samples (blood and urine) are resampled from the existing data sets of large cohorts, i.e., “Families segregating Type II diabetes and related traits”<sup>68</sup> and “The St Thomas’ UK Adult Twin Registry,”<sup>69</sup> the latter including over 4800 Caucasian twin pairs with extensive bioclinical data. For the metabonomics studies, the focus is on samples likely to show marked (and to some extent predictable) changes, e.g., diabetic vs nondiabetic.

The research work plan within the metabonomics work package has initially established the technology such that high-throughput metabonomic analysis can be generated using NMR and UPLC-MS techniques. Because of the complementary nature of the analytical techniques (in terms of qualitative and quantitative power), a more extensive metabolite picture is obtained. Data obtained by these techniques allow statistical analysis to be performed using pattern recognition and multivariate data analysis. MolPAGE is an exploratory study, and quantitative values describing sensitivity and specificity requirements for the identification of biomarkers have not been defined in advance at this early stage. Metabonomic methodology has the capacity to provide substantial amounts of information about small molecule metabolites that are likely to be disturbed in metabolic disease and to produce patterns of metabolites that correlate with disease onset and progression, as well as disease-related patterns from other genomic technologies.

The general objective of the data collection and integrated data analysis of the MolPAGE project is to develop a data warehouse for storing data from metabonomics, transcriptomics, and proteomics experiments and sample annotation metadata in a consistent, well-structured and annotated form, linked to other relevant data resources, as shown in **Figure 2**. Eventually all the (anonymized) data will be made available via the Internet and linked to other relevant data resources, such as the microarray data repository ArrayExpress.<sup>70</sup>

## Concluding Remarks

Metabonomic methodology is complementary to other functional genomic approaches. In a clinical context, metabonomics offers the advantage of using readily available biofluids, i.e., urine and plasma. This approach expands the number of metabolic markers available to the practitioner by an order of magnitude (otherwise not available through routine assays). The metabonomic untargeted approach enables characterization of early markers of disease and prognosis, as well as drug treatment efficacy<sup>71</sup> and eventually personalized health care. Unlike the other “-omics” (genomics, transcriptomics, proteomics), metabonomics does not focus on genes or their expression products (transcripts/proteins), but on metabolites. For this reason, metabonomics represents a rather exhaustive metabolic phenotyping technology, which will help in understanding metabolic diseases in general and diabetes in particular. The amount of information generated with metabonomics does not in itself a guarantee for the discovery of either a single predictive biomarker or a panel. However, the amount



**Figure 2.** Overview of the main components of MolPAGE scientific program.<sup>67</sup>





