



## A GC–MS metabolic profiling study of plasma samples from mice on low- and high-fat diets<sup>☆</sup>

Konstantina Spagou<sup>a</sup>, Georgios Theodoridis<sup>b</sup>, Ian Wilson<sup>c</sup>, Nikolaos Raikos<sup>a</sup>, Peter Greaves<sup>d,1</sup>, Richard Edwards<sup>d</sup>, Barbara Nolan<sup>d</sup>, Maria I. Klapa<sup>e,f,g,\*</sup>

<sup>a</sup> Laboratory of Forensic Medicine and Toxicology, School of Medicine, Aristotle University, Thessaloniki 54124, Greece

<sup>b</sup> Department of Chemistry, Aristotle University, Thessaloniki 54124, Greece

<sup>c</sup> Department of Clinical Pharmacology, Drug Metabolism and Pharmacokinetics, Astrazeneca, Alderley Park, Cheshire SK10 4TG, UK

<sup>d</sup> MRC Toxicology Unit, University of Leicester, LE1 7RH, UK

<sup>e</sup> Metabolic Engineering and Systems Biology Laboratory, Institute of Chemical Engineering and High-Temperature Chemical Processes (ICE-HT), Foundation for Research and Technology-Hellas (FORTH), Patras 26504, Greece

<sup>f</sup> Department of Chemical & Biomolecular Engineering, University of Maryland, College Park, MD 20742, USA

<sup>g</sup> Department of Bioengineering, University of Maryland, College Park, MD 20742, USA

### ARTICLE INFO

#### Article history:

Received 31 August 2010

Accepted 24 January 2011

Available online 4 February 2011

#### Keywords:

GC–MS metabolomics

Metabonomics

Obesity

Hyperglycemia

Quantitative systems biology

Multivariate statistical analysis

### ABSTRACT

Metabolic profiling of biofluids, based on the quantitative analysis of the concentration profile of their free low molecular mass metabolites, has been playing increasing role employed as a means to gain understanding of the progression of metabolic disorders, including obesity. Chromatographic methods coupled with mass spectrometry have been established as a strategy for metabolic profiling. Among these, GC–MS, targeting mainly the primary metabolism intermediates, offers high sensitivity, good peak resolution and extensive databases. However, the derivatization step required for many involatile metabolites necessitates specific data validation, normalization and analysis protocols to ensure accurate and reproducible performance. In this study, the GC–MS metabolic profiles of plasma samples from mice maintained on 12- or 15-month long low (10 kcal%) or high (60 kcal%) fat diets were obtained. The profiles of the trimethylsilyl(TMS)-methoxime(MeOx) derivatives of the free polar metabolites were acquired through GC–(ion trap)MS, using [U-<sup>13</sup>C]-glucose as the internal standard. After the application of a recently developed data correction and normalization/filtering protocol for GC–MS metabolomic datasets, the profiles of 48 out of the 77 detected metabolites were used in multivariate statistical analysis. Data mining suggested a decrease in the activity of the energy metabolism with age. In addition, the metabolic profiles indicated the presence of subpopulations with different physiology within the high- and low-fat diet mice, which correlated well with the difference in body weight among the animals and current knowledge about hyperglycemic conditions.

© 2011 Elsevier B.V. All rights reserved.

### 1. Introduction

Obesity is a major and rapidly expanding medical problem for modern western societies [1]. Understanding the progression of the disease and the molecular mechanisms characterizing its effect on the physiology of the affected individual and its connection with other hormonal dysfunctions is of great importance for identifying ways to reverse and/or prevent its distressing consequences [2]. To

this extent, analysis of the molecular composition of biofluids at the protein and metabolic levels has served in monitoring molecular changes that characterize the effect of high-fat diet on the onset and progression of obesity in animal studies and the human [3–5]. However, identifying specific and sensitive markers of diet-related effects in biofluids, the latter being the mirror for physiological changes occurring in the body, is not a trivial task. In the systems biology era, the quest for accurate and sensitive markers of complex biological phenomena has shifted towards the identification of characteristic multi-component molecular profiles rather than depending on single molecules [6,7]. Therefore, the development of validated methods for the accurate metabolite profiling of biofluids is important, particularly for determining the progression of metabolic dysfunctions, including obesity [8,9].

Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) coupled with a chromatographic method are

<sup>☆</sup> This paper is part of the special issue “Enhancement of Analysis by Analytical Derivatization”, Jack Rosenfeld (Guest Editor).

\* Corresponding author at: FORTH/ICE-HT, Stadiou, St. Platani, Rio-Patras, GR-26504, Greece. Tel.: +30 2610 965249; fax: +30 2610 965223.

E-mail address: [mklapa@iceht.forth.gr](mailto:mklapa@iceht.forth.gr) (M.I. Klapa).

<sup>1</sup> Current address: Department of Cancer Studies and Molecular Medicine, University of Leicester, UK.

the analytical techniques, which have been mainly used for high-throughput metabolic profiling analysis (metabo-l- or -n- omics), either separately or in combination for a large number of applications and biological systems [10–15]. The advantages of MS compared to NMR remain (a) the generally higher sensitivity of MS, which enables the use of smaller quantities of biological material [12,16,17], and (b) the potential for the analysis of a larger number of metabolites in a single run [16,17]. Among the “hyphenated” MS methods for metabolomic analysis, gas chromatography (GC)–MS with electron ionization remains a technique of choice for the metabolic profiling of polar intermediates of primary metabolism due to its high sensitivity, excellent resolution compared to LC due to a better discrimination of the compounds in the gas than in the liquid phase, capability for unknown compound identification based on the fragmentation pattern and well-established analyte databases for metabolite identification [12,18]. These, and other advantages [13], render GC–MS a useful and lasting tool for metabolomic-based molecular diagnosis. However, GC–MS-based metabolomics requires the derivatization of the extracted metabolites into volatile and thermally stable compounds. Due to the derivatization step, specific data validation, correction/filtering and normalization procedures are necessary to ensure comparability between profiles and to avoid assigning biological significance to experimental biases, which are due mainly to incomplete derivatization and the formation of multiple derivatives for some compounds [18,19]. Lacking appropriate GC–MS metabolomic data correction methodology from derivatization biases, GC–MS metabolomics has not been the technique of choice for the analysis of biological fluids in the context of obesity and high-fat diet research. Thus, there is a considerably limited number of relevant GC–MS metabolomic analyses for any biological fluid compared to the LC–MS and NMR studies. GC–MS has often been used for the metabolic profiling of plant or animal tissues, in which more complex matrices are encountered, but usually of larger sample volume and thus metabolite extract than in the analysis of biological fluids, especially blood plasma. In the case of small plasma sample volumes, the better discriminatory power of GC compared to the LC for some small metabolites, which are key intermediates and markers of activity of specific pathways, is of importance. In addition, biological fluids being the mirror of activity of many tissues, the higher sensitivity of GC–MS compared to LC–MS could allow for the quantification of metabolites that are in small quantities and could potentially indicate subtle physiological differences between samples that could have been unobservable in the LC–MS profile. For these advantages to be available, the GC–MS metabolic profile has to be appropriately corrected from experimental biases.

In this study, we applied GC–MS metabolic profiling for the analysis of plasma samples from mice maintained on 12- or 15-month long low- or high-fat diets, in order to characterize metabolic differences between the groups. The GC–MS metabolic profiles were obtained using an ion-trap MS and a recently developed data correction and normalization methodology for derivatization biases [18,19] was applied on the acquired dataset to ensure accurate and validated performance. Here, the application of this data normalization method is described in detail to provide a reference for its application in other biological fluids too. The corrected metabolic profiles were analyzed using multivariate statistical analysis, the results of which were interpreted in the context of the currently known physiological information about the effect of high- and low-fat diets. This enabled the identification of subgroups of animals, which phenomenologically may have had a different than “expected” metabolic physiology, but the careful analysis of the multi-compound profiles indicated special characteristics that clustered them together. Our objective is to show that if appropriately corrected, GC–MS metabolomic data could be a valuable resource either alone or in combination with LC–MS and NMR data

**Table 1**

The body weight (BW) of the mice in this study on the day of the blood sample acquisition.

Animal ID	Diet	BW (g)	Mean BW $\pm$ SD
Diet duration: 12 months			
LF3	Low-fat	38	
HF3	High-fat	57	
HF8	High-fat	45	
HF9	High-fat	55	
HF11	High-fat	60	53.6 $\pm$ 7.3
HF12	High-fat	63	
HF13	High-fat	51	
HF14	High-fat	44	
Diet duration: 15 months			
LF1	Low-fat	42	
LF2	Low-fat	47	41 $\pm$ 6.6
LF4	Low-fat	34	
HF1	High-fat	58	
HF2	High-fat	67	
HF4	High-fat	58	
HF5	High-fat	52	
HF6	High-fat	60	55.4 $\pm$ 6.5
HF7	High-fat	51	
HF15	High-fat	48	
HF16	High-fat	49	

Low-fat diet: 10 kcal%.

High-fat diet: 60 kcal%.

SD, standard deviation

for the analysis of the biological fluid composition in the context of obesity research.

## 2. Materials and methods

### 2.1. Experimental design and setup

Nineteen (19) C57Bl6J mice obtained from Harlan UK Ltd. (Bicester, UK) were used in this study separated into two different groups based on their diet (see Fig. 1). Specifically, four (4) mice were fed low-fat purified rodent diet with 10 kcal% fat (D12450B) and fifteen (15) a similar diet but containing 60 kcal% fat (D12492), both purchased from Research Diets Inc., New Brunswick, NJ, USA. The experiment was carried out under project license PPL 40/2527 granted by the UK Home Office and in accordance with the Home Office regulations and the protocol was reviewed by the Leicester University Local Ethical Committee for Animal Experimentation.

Seven (7) animals from the group of high-fat diet and one (1) animal from the group of low-fat diet were sacrificed 12 months after commencing the diet, while the rest (i.e. eight (8) from the high-fat diet group and three (3) from the low-fat diet group) were sacrificed 15 months after commencing the diet (according to the standards of the international statutes on pain minimization (86/609/EEC)), and a full autopsy was conducted. The weights of the animals at the autopsy are shown in Table 1. Blood samples were obtained by tail puncture. Plasma was isolated after centrifugation of the blood samples at 1700  $\times$  g and 4 °C during the first hour after blood acquisition. The plasma samples were stored at –80 °C pending analysis.

### 2.2. Metabolic profile acquisition and normalization

[U-<sup>13</sup>C]-glucose (Cambridge Isotope Laboratories, Cambridge, MA, USA) was added to the plasma samples as the internal standard at a concentration of 0.6  $\mu$ L/ $\mu$ L of plasma and the samples were subsequently dried in vacuum at room temperature in a pre-cooled rotor. In light of their small volume, the duration of the vacuum drying was short for all samples, while careful monitoring of the drying process limited significantly the probability for

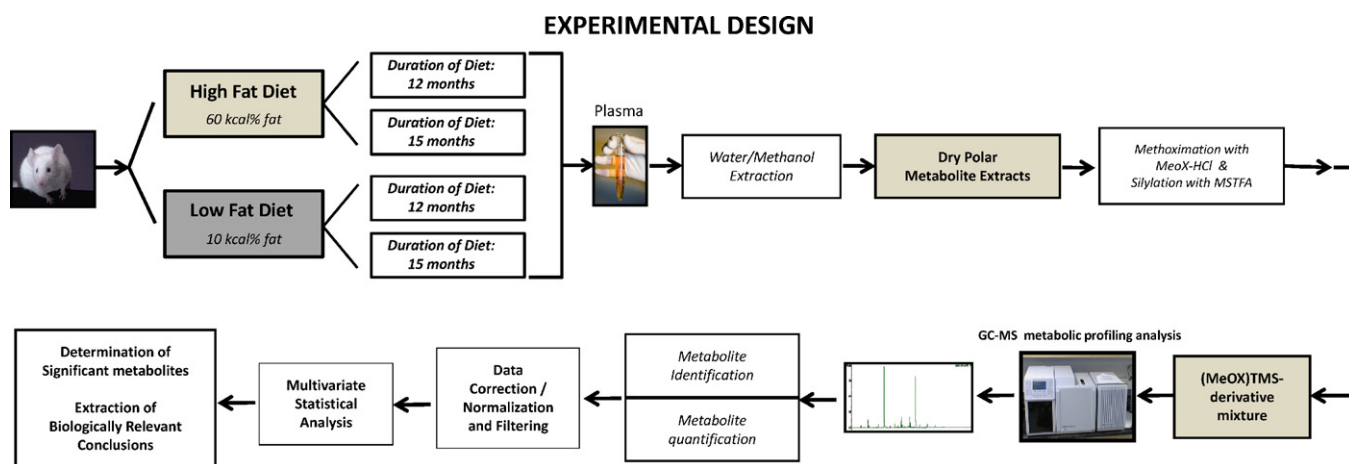


Fig. 1. Schematic of the experimental design.

metabolite evaporation. Cold HPLC-grade methanol was added to the dried plasma samples at the volumes shown in [Supplementary File 1](#); the samples were placed in a 70 °C water-bath for 15 min and a volume of water equal to that of the methanol was then added. After centrifugation at 1500 × g for 10 min, the supernatants were concentrated in vacuum. The dried polar extracts were derivatized to their (MeOx)TMS-derivatives through reaction with 20 mg/mL methoxyamine hydrochloride (Alfa Aesar, UK) solution in pyridine for 90 min, followed by reaction with N-methyl-trimethylsilyl-trifluoroacetamide (MSTFA) (Alfa Aesar, UK) at room temperature for at least 6 h (see [18,19] and further explanation in Section 3); the volumes of methoxyamine hydrochloride and MSTFA used for each sample are shown in [Supplementary File 1](#). The metabolic profiles of the derivatized samples were acquired using the Saturn 2200 gas chromatograph–(ion trap) mass spectrometer (Varian Inc., CA, USA) at 1:10 split ratio (GC column: Phenomenex #7HG-G004-11-B, USA). The peak identification and quantification were carried out as described in [19]. The raw metabolomic dataset comprised 77 peaks, each of which corresponded to a compound of known chemical category according to the category definition in [18,19], and had been detected in at least one of the acquired metabolic profiles.

The relative areas of all detected peaks (RPAs – relative peak areas) were estimated from their normalization with the 323 marker ion peak area of the second derivative (MeOx2) of the internal standard [U-<sup>13</sup>C]-glucose. Moreover, the recently developed data validation, normalization and correction methodology for GC–MS metabolomic datasets [19] was applied to account for the derivatization biases [18,19]. First, we verified same GC–MS operational conditions during the acquisition of all metabolic profiles based on the criterion described in [18,19], using the ratio of the 323 marker ion peak areas of the two derivatives of the internal standard [U-<sup>13</sup>C]-glucose. According to this criterion, four metabolic profiles were excluded from further analysis (shown in orange background in [Supplementary File 2A](#)). It has to be noted that according to measurements in our laboratory using glucose standards and mouse plasma samples from the same batch as those used in the present study (i.e. with similar concentration of intracellular glucose), we observed (as it is the case with other biological matrices too) that at these internal glucose concentrations, the contribution of its 323 (M+4) ion peak area to the measured 323 ion peak area, which is used as characteristic of the internal standard, is almost non-existent and cannot affect the accuracy of the internal standard measurement used for data validation and normalization. Second, we calculated the cumulative (effective) RPA for the amine-group containing metabolites for which more than one derivative

peaks were detected in the metabolic profiles as the weighted summation of all their derivative RPAs, according to the data correction strategy described in [18,19]. In this case, only alanine was identified with two derivative peaks. The metabolic profile of sample HF16 (a mouse on the 15-month high-fat diet), which was acquired at seven different silylation times equal to or longer than 6 h (see [Supplementary File 2A](#)), was used for the estimation of the weight coefficients for the two derivative RPAs of alanine based on the algorithm described in [19]. The amine-group containing metabolites, for which only one derivative was detected in at least one of the metabolic profiles (i.e. ethanolamine, lysine, glutamate, glycine, serine and valine), but for which more than one derivative is known, were also considered; most often they are filtered out at a later step (in this case this happened for all but lysine), because of a high coefficient of variation for their RPAs between injections of the same sample (see below). Further, (a) the smallest of the two MeOx peaks of the known ketone-group containing metabolites (see [19]), (b) the peaks corresponding to unknown amine-group containing metabolites (see [19]), (c) the peaks (i) that were identified as artifacts due mainly to column bleeding, or (ii) with significant carry over, or (iii) that were inconsistently detected among samples and/or injections of the same sample, or (iv) with significant mean coefficient of variation between injections over all samples, were filtered out of the analysis. After this normalization and filtering step, the RPA profiles that were used for further analysis included 50 metabolites (shown in [Supplementary File 2B](#)). Subsequently, the arithmetic mean of all metabolic profiles acquired for each mouse was estimated and these 19 mean metabolic profiles (provided in [Supplementary File 2](#)) were used in the TM4 MeV (v4.6) omic data analysis software [20] with an 80% cut-off. In this way, two (2) additional metabolites were excluded from further analysis, the unknown\_1328 and 4-hydroxybutyrate. In the final 48-metabolite profiles, any missing RPAs were imputed using the *k*-nearest neighbor's algorithm [21], as implemented in the TM4 MeV software.

### 2.3. Metabolomic data analysis

The multivariate statistical analyses applied to the 48-metabolite profiles were based on both the non-standardized and the standardized metabolite RPAs. Specifically, the standardized RPA of metabolite M in the *j*-th metabolic profile,  $stRPA_{Mj}$ , is equal to:

$$stRPA_{Mj} = \frac{RPA_{Mj} - \overline{RPA_M}}{SD_{RPA_M}} \quad (1)$$

where  $RPA_{M_j}$ ,  $\overline{RPA_M}$ ,  $SD_{RPA_M}$  are, respectively, the RPA of metabolite M in the  $j$ -th metabolic profile, the mean RPA of metabolite M over all metabolic profiles and the standard deviation of the RPA of metabolite M over all metabolic profiles. The use of standardized relative peak areas in these analyses was required due to the large variance in the order of magnitude of the RPAs among the different metabolites in the same metabolic profile. When non-standardized values are used, this variance biases the results in accordance mainly with the variation in the concentrations of the higher-order of magnitude metabolites, as it will be shown in Section 3.

The “omic” data analysis software TM4 MeV (v4.6) [20] was used to mine the metabolomic dataset. Specifically, hierarchical clustering (HCL) using Euclidean, Manhattan or Pearson correlation distance metrics was used to cluster the samples or the metabolites based on their profiles. Principal component analysis (PCA) was used to visualize whether the various plasma samples could be differentiated based on their metabolic profiles. The metabolites, the concentration of which was significantly higher or lower in a set of plasma samples compared to another, will be referred to as positively or negatively, respectively, significant metabolites of the particular comparison. The significant metabolites were identified using unpaired Significance Analysis of Microarrays (SAM) [22].

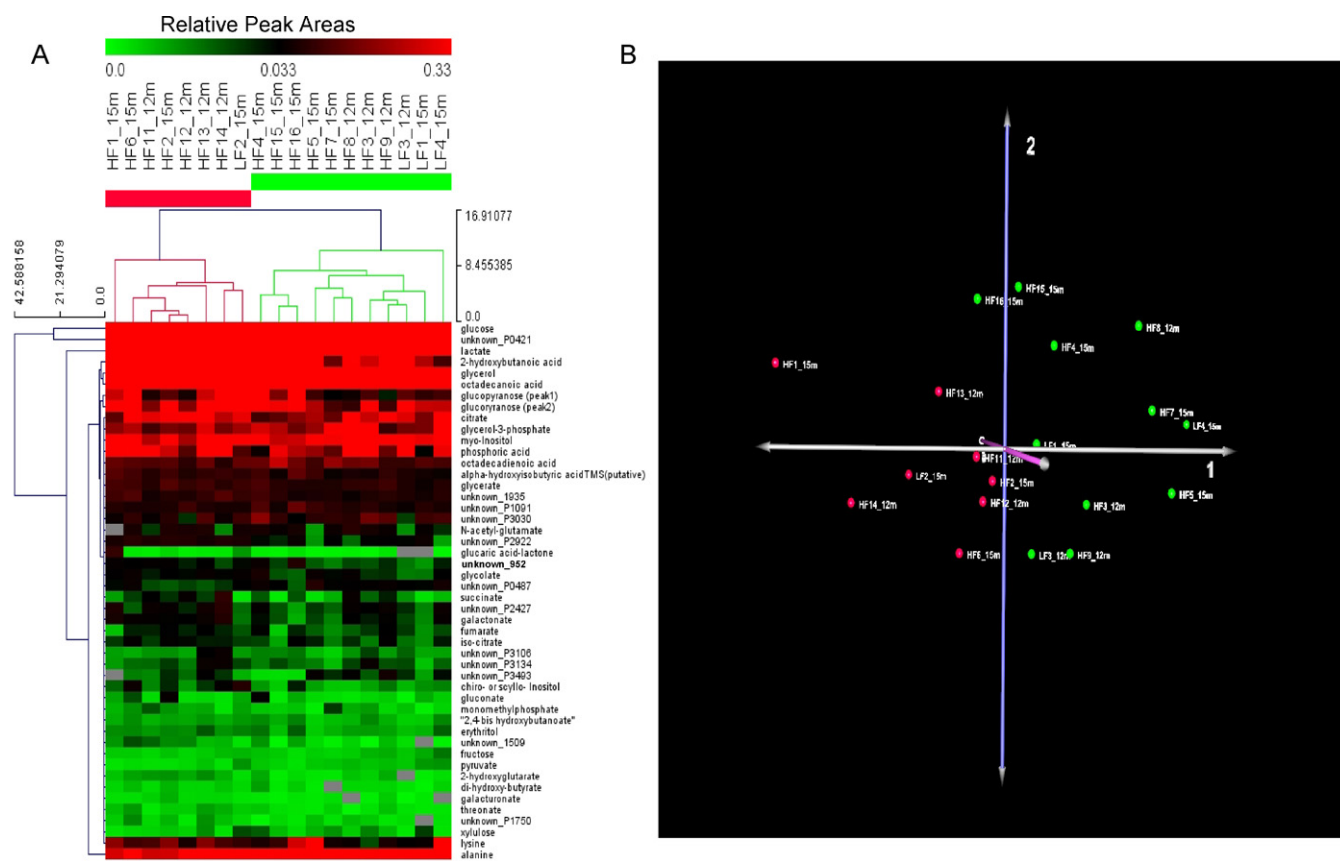
### 3. Results and discussion

#### 3.1. Correction strategy for the derivatization biases

In the case of a raw GC–MS metabolomic dataset, two types of biases are encountered and need to be corrected before any analysis

to extract biologically relevant conclusions can be applied [18]. The first type is common among all analytical techniques and refers to variations among metabolic profiles due to e.g. differences in the original sample size, or systematic biases introduced at the various steps of the multi-step procedure, which affect equally all metabolites in the original sample. This type of bias is corrected with the use of (an) internal standard(s) spiked into all samples at the same concentration with respect to the size of the sample. The compound(s) used as internal standard(s) should not be among the endogenous metabolites of the analyzed biological sample. The internal standard concentration should be selected similar to the mean metabolite concentration expected to be measured in the analyzed biological sample to avoid dividing the metabolite peak areas with a very large or a very small number. Ribitol has been an internal standard of choice in GC–MS metabolomics [18,23,24], however, ribitol was detected among the metabolites potentially present in the mouse blood plasma samples. Therefore,  $[U-^{13}C]$ -glucose was used as the internal standard in this study, as a less expensive and least interfering with the measurements labelled compound.

The second type of biases found in a GC–MS metabolomic dataset result from the derivatization step that is required for the metabolites to become volatile and thermally stable to be measured by the GC. These biases affect the various metabolites in a different way. They may originate from the acquisition of the various metabolic profiles under different GC–MS equipment conditions and/or at different stages of derivatization; if these biases are not corrected, there is a risk of assigning biological significance to differences due to incomplete derivatization and/or matrix effects (see [18,19]). An additional source of these biases for some metabo-



**Fig. 2.** (A) HCL (Euclidean distance for both samples and metabolites) and (B) PCA of the profiles of the non-standardized metabolic profiles of the 19 plasma samples. The red and green color-code refers to the same profiles in the two analyses. PC1, PC2 and PC3 refer to the % profile variation from the original experimental space carried by principal components 1, 2 and 3, respectively. The discriminatory metabolites between the red and the green groups are shown in Table 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

lites could be the formation of multiple derivatives, whose relative concentration with respect to the original metabolite concentration may vary with the derivatization time. To avoid introducing any related biases to the metabolic profiling analysis, the derivatization time at which the metabolic profiles are acquired, and the derivative peak area, which will be used in the multivariate statistical analysis as proportional to the concentration of the original metabolite, have to be carefully selected (see [18,19]). For the two-step derivatization procedure used in this study, Kanani and Klapa [18,19] divide the metabolites, which react with the derivatization agents, into three chemical categories: the first comprises the metabolites that contain neither ketone- nor amine-groups, the second comprises the metabolites which contain ketone- but no amine-groups and the third those metabolites which contain amine-groups. The first category forms only one derivative per metabolite. Due to the methoximation step, the second category forms two isomeric derivatives per metabolite, whose concentration ratio is constant based on the underlying chemistry. Thus, barring changes in the GC–MS equipment conditions, the ratio of the two isomeric derivative peak areas should remain constant among all acquired metabolic profiles [18]. This robust quality control criterion was used in the present study to ensure constant

GC–MS equipment conditions among all compared metabolic profiles. Based on the ratio of the 323 marker ion peak areas of the two derivatives of the internal standard [U-<sup>13</sup>C]-glucose, four metabolic profiles were excluded from further analysis (shown in orange background in [Supplementary File 2A](#)). Moreover, since each of the derivative peak areas of the second category of metabolites is proportional to the concentration of the original metabolite, only one of the two in each of the remaining metabolic profiles is used in multivariate statistical analysis to avoid mathematical biases (see also [19]); the smallest derivative peak area is usually excluded as prone to a higher signal to noise ratio.

The third category of metabolites may sequentially form more than one silylation derivative; thus, unless the silylation of these metabolites has fully completed and only the final derivative is present, more than one peak of varying RPA, depending on the stage of silylation, are to be detected in the metabolic profile. It has been shown [18] that complete silylation of all the third category metabolites may require more than 20 h; so long silylation times render the procedure impractical, while compound degradation effects may also become an issue. In [19], a strategy for the correction of the GC–MS metabolic profiles from biases originating from the presence of multiple derivatives for the third category

**Table 2**

The positively significant metabolites in the set of profiles colored “red” compared to the set of profiles colored “green” in the HCL tree of [Fig. 2A](#) based on unpaired SAM analysis.

	Positively significant metabolite no.	Positively significant metabolite name	Order of significance
	1.	unknown_P2922	↓
	2.	glucose	
	3.	unknown_P0421	
	4.	2-hydroxyglutarate	
	5.	unknown_1509	
	6.	threonate	
	7.	chiro- or scyllo-inositol	
	8.	galactonate	
	9.	glycerate	
	10.	unknown_1935	
	11.	glucopyranose (peak 1)	
	12.	glucopyranose (peak 2)	
	13.	unknown_P1750	
	14.	isocitrate	
	15.	unknown_P2427	
	16.	succinate	
	17.	alpha-hydroxyisobutyric acid (putative)	
	18.	galacturonate	
	19.	erythritol	

Significance threshold value d1 corresponds to zero False Discovery Rate (FDR)-median and 0.03% FDR-90-th percentile (i.e. 0.25 false positive metabolites). Significance threshold value d2 corresponds to 0.04% False Discovery Rate (FDR)-median (i.e. 0.75 false positive metabolites) and 0.11% FDR-90-th percentile (i.e. 2 false positive metabolites). The metabolites are shown in decreasing order of significance. The corresponding SAM graph is shown in [Supplementary File 3A](#).

metabolites was introduced. The only requirement for this strategy to be correctly applied is that all metabolic profiles are acquired at silylation times at which all the original third category metabolites in the sample have been fully converted into at least one of their derivatives, while the silylation reaction may still be proceeding. It is thus important for the experimentalist to determine the minimum silylation time,  $T_M$ , for this condition to be valid for a particular biological sample [see 19]. At these silylation times, it is the weighted sum of all derivative RPAs (i.e. the cumulative or effective RPA) of a third category metabolite that is proportional to its concentration in the original sample and can be used as its representative in any following statistical analysis to extract biologically relevant conclusions. Kanani and Klapa [19] introduced a methodology for the estimation of the weights of the derivative RPAs in this weighted sum. For the methodology to be applied, the metabolic profile of one sample from the experimental set should have been measured at multiple silylation times after  $T_M$ .

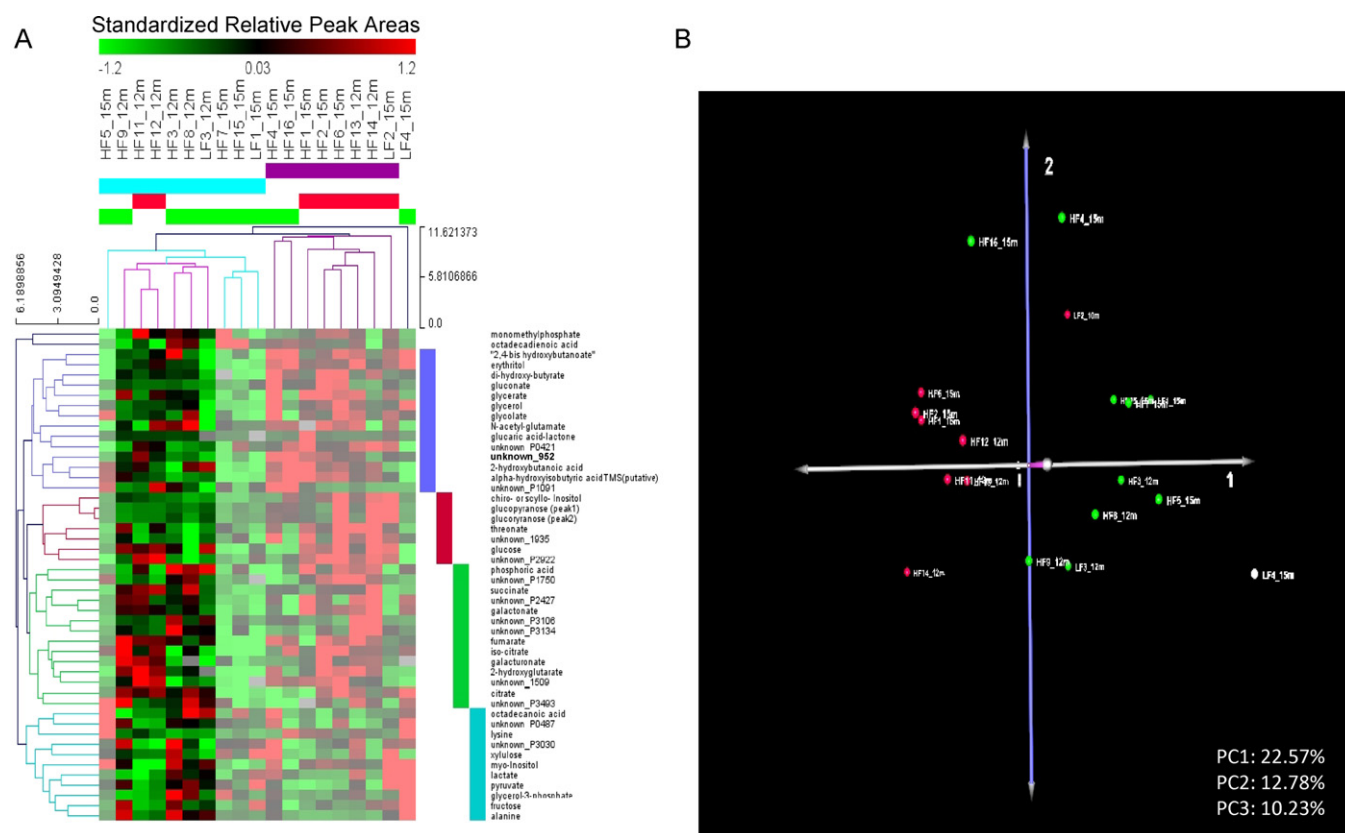
It has been shown (see [18]; also in one of the blood samples used in this study (data not shown)) that after 6 h, the silylation of the compounds without amine-groups will have been completed and the original amine-group containing metabolites should have been fully converted into at least one of their derivatives. Therefore, all the metabolic profiles of the present study were acquired at least twice at silylation times longer than 6 h. Only alanine was the third category metabolite with more than one derivative peak detected in the metabolic profiles. The weights of the two alanine derivative RPAs for the estimation of its cumulative RPA in all metabolic profiles were calculated based on the metabolic profile of sample HF16, which was measured at seven silylation times equal to or longer than 6 h. The two derivative RPAs of alanine in each metabolic profile were then replaced by its cumulative RPA. Since the peak areas of derivatives attributed to third category metabo-

lites of currently unknown identity cannot be combined based on the previously mentioned data correction strategy, they were filtered out of further analysis to avoid introducing any undesired biases. After these data validation, normalization and correction steps were performed (see also Section 2), the metabolomic dataset, now corrected from derivatization biases, was further filtered from peaks that were inconsistently detected and/or with a high signal to noise ratio and/or with significant carry over and/or with high average coefficient of variation between injections over all the samples, leading to the final set of 50-metabolite profiles (see Supplementary File 2B). The latter was used with an 80% cut-off in multivariate statistical analysis to extract biologically relevant conclusions as described below.

### 3.2. Multivariate statistical analysis of the metabolic profiles

This study comprised four groups of mice differing in diet (low-fat vs. high-fat diet) and duration of diet (12 vs. 15 months). According to Table 1, there was no statistical difference between the mean BWs of the 12-month and 15-month high-fat diet groups. This was also true for the low-fat diet animals. However, in this group, the plasma sample of only one 12-month diet mouse was available. Moreover, one (LF4) of the three 15-month low-fat diet mice was of significantly lower BW than the other two, even lower than the 12-month low-fat diet mouse. The mean BW of the high-fat diet mouse groups was higher than the mean BW of the same duration low-fat diet groups. However, there was an 11–13% variation in the BW of the mice within the high-fat diet groups; this macroscopic physiological parameter implies diversity in the metabolic physiology of the animals within these groups.

Indeed, hierarchical clustering (HCL) analysis of the non-standardized metabolomic dataset identified two main clusters,



**Fig. 3.** (A) HCL (Euclidean distance for both profiles and metabolites) and (B) PCA of the profiles of the metabolites identified as significant from SAM for the significance threshold value  $d1 = 1.91$  (see Fig. 2 and Table 2). The red and green color-code refers to the same profiles assigned the particular color in the analyses shown in Fig. 2. PC1, PC2 and PC3 are defined as in Fig. 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

**Table 3**

The positively significant metabolites in the set of profiles colored “purple” compared to the set of profiles colored “light-blue” in the HCL tree of Fig. 3A, based on unpaired SAM analysis (standardized RPA values).




	Positively significant metabolite no.	Positively significant metabolite name	Order of significance
	1.	erythritol	
	2.	alpha-hydroxyisobutyric acid (putative)	
	3.	di-hydroxybutyrate	
	4.	unknown_1935	
	5.	unknown_P0421	
	6.	chiro- or scyllo-inositol	
	7.	threonate	
	8.	glycerol	
	9.	2,4-bis hydroxybutanoate	
	10.	glucopyranose (peak 1)	
	11.	glucopyranose (peak 2)	
	12.	unknown_952	
	13.	glycerate	
	14.	gluconate	
	15.	unknown_P2922	
	16.	2-hydroxy-butanoic acid	
	17.	isocitrate	
	18.	unknown_1509	
	19.	glycolate	
	20.	2-hydroxyglutarate	
	21.	unknown_P1750	
	22.	galactonate	
	23.	unknown_P3106	
	24.	glucaric acid-lactone	

Significance threshold value d1 corresponds to zero False Discovery Rate (FDR)-median and 0.05% FDR-90-th percentile (i.e. 1 false positive metabolite). Significance threshold value d2 corresponds to 0.023% False Discovery Rate (FDR)-median (i.e. 0.66 false positive metabolites) and 0.125% FDR-90-th percentile (i.e. 3 false positive metabolites). The metabolites are shown in decreasing order of significance. The corresponding SAM graph is shown in [Supplementary File 3B](#).

which comprised both high-fat and low-fat diet mice of both diet durations. The two clusters are shown as the red and green branches of the HCL tree in Fig. 2A; using the same color scheme in principal component analysis (PCA) (see Fig. 2B), the “green” and “red” metabolic profiles are observed on the positive and negative, respectively, sides of principal component 1 (PC1), which carries 69% of the variance among the profiles in the original experimental space. Significance Analysis for Microarrays (SAM) indicated 19 metabolites whose increase in concentration in the “red” compared to the “green” group was discriminatory for this separation. The positively significant metabolites are shown in Table 2. While two of the most discriminatory metabolites are of yet unknown identity, i.e. the unknown\_P2922 (first) and unknown.P0421 (third), the second most discriminatory metabolite, whose concentration was significantly increased in the “red” compared to the “green”

group, was glucose (its two detected cyclic forms, glucopyranose (peak 1) and glucopyranose (peak 2) are also among the positively significant metabolites). Interestingly, the sixth most discriminatory metabolite is scyllo-inositol. Actually, the measured peak area could in theory correspond to both scyllo- and chiro-inositol, because these two stereoisomers of inositol cannot be distinguished by GC-MS [25]. However, scyllo-inositol is the only inositol stereoisomer apart from myo-inositol whose increase in concentration has been implicated with hyperglycaemic conditions [26]. The presence of isocitrate, succinate and 2-hydroxyglutarate, the latter being a direct product of alpha-ketoglutarate, among the positively significant metabolites, suggests a higher tricarboxylic acid (TCA) cycle activity in the “red” compared to the “green” mouse group. Glycerate, galactonate and threonate are also among the positively significant metabolites.

**Table 4**  
 (A) The positively significant metabolites in the set of the “red” profiles in Fig. 3B compared to the set of the “green” profiles except LF4. (B) The negatively significant metabolites in the profiles of the 15-month diet mice, except mice LF4 and HF5, compared to the profiles of the 12-month diet mice, based on unpaired SAM analysis (standardized RPA values).

A.	Positively Significant Metabolite no	Positively Significant Metabolite name	B.	Negatively Significant Metabolite no	Negatively Significant Metabolite name	Order of Significance
 $d1 = 1.47$	1.	unknown_P2922	 $d2 = 0.70$	1.	succinate	
	2.	2-hydroxyglutarate		2.	unknown_P3493	
	3.	galactonate		3.	unknown_P3134	
	4.	unknown_1509		4.	citrate	
	5.	galacturonate		5.	unknown_P3106	
	6.	threonate		6.	unknown_P2427	
	7.	glucose		7.	fumarate	
	8.	unknown_P2427		8.	galactonate	
	9.	unknown_P0421		9.	alanine	
	10.	unknown_P1750		10.	phosphoric acid	
	11.	succinate		11.	2-hydroxyglutarate	
	12.	isocitrate				

(A) Significance threshold value  $d1$  corresponds to zero False Discovery Rate (FDR)-median and 0.06% FDR-90-th percentile (i.e. 0.75 false positive metabolite).

(B) Significance threshold value  $d2$  corresponds to 0.16% False Discovery Rate (FDR)-median (i.e. 1.75 false positive metabolites). Significance threshold value  $d3$  corresponds to zero FDR-median.

The metabolites are shown in decreasing order of significance. The corresponding SAM graphs are shown in Supplementary File 3C and D.

It is important to discuss the members of the “red” metabolic profile group. These comprise animals from both the 12-month and the 15-month high-fat diet mice and even one low-fat diet mouse, LF2. The latter was indeed the heaviest among the low-fat diet mice, with a BW larger than two of the 12-month high-fat diet mice of this study. LF2 was measured with a very high concentration of plasma glucose (third highest) and the highest concentration of scyllo-inositol (see Supplementary File 2B) among all animals of the study. Moreover, the “red” metabolic profile group contains the two and three, respectively, heaviest 12-month and 15-month high-fat diet mice (i.e. HF11–12 and HF1, HF2, HF6). However, mouse HF4, which had the same weight as mouse HF1, is clustered in the “green” group. It is of interest to note that the “red” plasma metabolic profile group also comprises the profiles of the 12-month high-fat diet mice HF13 and HF14. This would not have been directly apparent from the BW measurement table, as these two mice are among the lightest of the high-fat diet animals; actually, HF14 was the lightest of all, with a BW similar to the low-fat diet mice of this study. However, HF14 was observed with the highest among all mice concentration of glucose, succinate and galactonate (unknowns P2427, P3106, P3134 too) in its blood plasma (see Supplementary File 2B).

Within the “green” metabolic profile group, it is of interest to discuss the plasma metabolic profile of the 15-month low-fat diet mouse LF4, whose BW was very low at the time of sample acquisition. As it can be observed in Fig. 2A, its profile “branches out” of the other “green” profiles in the HCL tree. LF4 was measured with the lowest concentration of the three most discriminatory metabolites between the “red” and “green” groups (i.e. unknown\_2922, glucose and unknown\_P0421) and 2-hydroxybutanoate, but the highest of alanine, fructose, pyruvate, citrate, glycerol-3-phosphate

(its concentration was dramatically higher than in any other plasma profile) and glycerol (its concentration was much higher than that seen in any other low-fat diet mouse plasma sample) and the second highest of lysine. This profile is consistent with active lipolysis (based on the increased concentration of glycerol-3-phosphate and glycerol) and proteolysis (based on the increased concentration of alanine, lysine and pyruvate) in LF4, as a consequence of the prolonged fasting. Such metabolic state could justify its significantly low BW.

Taking into consideration that multivariate statistical analysis on the non-standardized metabolomic values cannot reveal some interesting features of the dataset due to the large difference in the order of magnitude between the metabolites, HCL analysis was also carried out on the standardized metabolic profiles (see Section 2). The HCL heat map (Euclidean distance) shown in Fig. 3A validates the unique metabolic profile of LF4 among the other plasma samples, as discussed earlier. The rest of the metabolic profiles are divided into two groups, which coincide to a great extent with the “green” and “red” groups described earlier, with a higher metabolic activity in the latter (currently assigned a purple color) than the former (currently assigned a light blue color) as observed from the heat map. All mice on a 15-month high-fat diet with a BW less than the mean BW of the 12-month high-fat diet group, except mouse HF16, clustered together in the “lower metabolic activity” cluster. The HF16 mouse profile indicates a higher concentration of ketone bodies in its plasma (see purple metabolite cluster on the right of the heat map), a pattern which is consistent with keto-acidosis and insulin resistance. The profiles of HF1, HF2, HF6, HF13 and HF14, all of which belong to the “purple” group, were consistent with the expected hyperglycaemic metabolic state, with high concentra-



tions of glucose, scyllo-inositol and TCA cycle activity intermediates (see the heat map of the red and green metabolites in these groups in Fig. 3A). Moreover, the profiles obtained for all of these mice, with the exception of mouse HF14, also imply active keto-acidosis as a result of hyperglycaemic conditions. SAM analysis between the light blue and the purple metabolic profiles of the plasma samples verified these observations indicating as most discriminatory the higher concentrations of erythritol, scyllo-inositol, the cyclic forms of glucose, ketone-bodies, glycerol, gluconate, threonate and glycolate in the purple compared to the light blue cluster (see Table 3).

The 3-D PCA graph of the standardized metabolic profiles can be observed in Fig. 3B. It is of importance to note that (a) all members of the previously mentioned “red” mouse group are observed on the negative side of PC1 except from mouse LF2, (b) despite the high variance between their metabolic profiles, all low-fat diet mice are observed on the same (positive) side of PC1, and (c) the 15-month mouse plasma samples are separated from the 12-month mouse plasma samples on PC2; the only exceptions (apart from mouse LF4) are the heaviest 12-month high-fat diet mouse, HF12, which appears on the positive side of PC2 and the 15-month high-fat diet mouse HF5, which had a particular profile as it can be observed in Fig. 3A. The list of significant metabolites that separate the two sides of PC1 (mouse LF4 was excluded from the analysis and HF16 was considered together with the profiles on the positive side of PC1) is shown in Table 4A, indicating the glucose concentration, the TCA cycle activity (through the measurements of succinate, isocitrate and 2-hydroxyglutarate) and the concentration of the acids threonate, galactonate and galacturonate, to be significantly higher in the profiles of the negative (“red”) side of PC1 in Fig. 3B. These observations imply a more severe hyperglycaemic state in the “red” group compared to the other mice, as it was shown with the non-standardized profiles too. The results of SAM analysis between the plasma metabolic profiles of the 12 month and 15 month diet mice, excluding LF4 and HF5 (Table 4B) indicated the energy metabolism (i.e. the TCA cycle activity, as observed through the measurements of the metabolites, succinate, citrate, fumarate, and 2-hydroxyglutarate, and the concentration of phosphoric acid) and the concentrations of plasma alanine and galactonate to be characteristically higher in the 12-month compared to the 15-month diet mice. When only the high-fat diet mice are considered in this comparison, only the first four metabolites in Table 4B are identified as significant, including succinate and citrate as the two known.

#### 4. Conclusions

While GC–MS-based metabolomics is a powerful technique for the analysis of biofluids in a number of applications, the required derivatization step introduces an additional set of biases, which affect each metabolite in a different way. The dataset has to be appropriately corrected and normalized before used to extract biologically relevant conclusions. In this study, we measured the GC–MS metabolic profiles of a set of plasma samples from mice maintained on 12- or 15-month long low- or high-fat diets. The application of a recently developed GC–MS metabolomic data validation and normalization/filtering protocol enabled the iden-

tification and filtering of non-consistently measured profiles and metabolites, ensuring a final metabolomic dataset free of derivatization biases. The final 48-metabolite profiles were analyzed using multivariate statistical analysis techniques. The acquired results indicated (a) increased energy metabolism activity in the 12- compared to the 15-month diet mice and (b) the presence of subpopulations of different metabolic physiology within the main four mouse groups, which correlated well with the observed difference in the BW of the animals. In the case of hyperglycaemic high-fat diet animals, the metabolic profile agreed well with current knowledge about the relevant metabolic physiology, indicating in most cases simultaneously active ketoacidosis and insulin resistance.

#### Acknowledgment

We would like to gratefully acknowledge the financial support of FORTH/ICE-HT to Dr. Klapa that funded all expenses related to the acquisition of the metabolic profiles.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jchromb.2011.01.028.

#### References

- [1] Y. Wang, M.A. Beydoun, L. Liang, B. Caballero, S.K. Kumanyika, *Obesity* (Silver Spring) 16 (2008) 2323.
- [2] T.T. Huang, T.A. Glass, *JAMA* 300 (2008) 1811.
- [3] P.G. Kopelman, *Nature* 404 (2000) 635.
- [4] M.E. Dumas, R.H. Barton, A. Toye, O. Cloarec, C. Blancher, A. Rothwell, J. Fearnside, R. Tatoud, V. Blanc, J.C. Lindon, S.C. Mitchell, E. Holmes, M.I. McCarthy, J. Scott, D. Gauguier, J.K. Nicholson, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 12511.
- [5] S.H. Kim, S.O. Yang, H.S. Kim, Y. Kim, T. Park, H.K. Choi, *Anal. Bioanal. Chem.* 395 (2009) 1117.
- [6] J.K. Nicholson, J.C. Lindon, E. Holmes, *Xenobiotica* 29 (1999) 1181.
- [7] M.I. Klapa, J. Quackenbush, *Biotechnol. Bioeng.* 84 (2003) 739.
- [8] J.K. Nicholson, I.D. Wilson, *Nat. Rev. Drug Discov.* 2 (2003) 668.
- [9] J.L. Griffin, A.W. Nicholls, *Pharmacogenomics* 7 (2006) 1095.
- [10] E.J. Want, B.F. Cravatt, G. Siuzdak, *ChemBiochem* 6 (2005) 1941.
- [11] J.C. Lindon, E. Holmes, J.K. Nicholson, *Pharm. Res.* 23 (2006) 1075.
- [12] K. Dettmer, P.A. Aronov, B.D. Hammock, *Mass Spectrom. Rev.* 26 (2007) 51.
- [13] E.M. Lenz, I.D. Wilson, *J. Proteome Res.* 6 (2007) 443.
- [14] K.K. Pasikanti, P.C. Ho, E.C. Chan, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* 871 (2008) 202.
- [15] G. Theodoridis, I.D. Wilson, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* 871 (2008) 141.
- [16] J. Kopka, *J. Biotechnol.* 124 (2006) 312.
- [17] J. Kopka, A. Fernie, W. Weckwerth, Y. Gibon, M. Stitt, *Genome Biol.* 5 (2004) 109.
- [18] H. Kanani, P.K. Chrysanthopoulos, M.I. Klapa, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* 871 (2008) 191.
- [19] H.H. Kanani, M.I. Klapa, *Metab. Eng.* 9 (2007) 39.
- [20] A.I. Saeed, V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturm, M. Snuffin, A. Rezzantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, J. Quackenbush, *Biotechniques* 34 (2003) 374.
- [21] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, R.B. Altman, *Bioinformatics* 17 (2001) 520.
- [22] V.G. Tusher, R. Tibshirani, G. Chu, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 5116.
- [23] U. Roessner, C. Wagner, J. Kopka, R.N. Trethewey, L. Willmitzer, *Plant J.* 23 (2000) 131.
- [24] B. Dutta, H. Kanani, J. Quackenbush, M.I. Klapa, *Biotechnol. Bioeng.* 102 (2009) 264.
- [25] S.K. Fisher, J.E. Novak, B.W. Agranoff, *J. Neurochem.* 82 (2002) 736.
- [26] D.A. Simmons, A.I. Winegrad, *Diabetologia* 32 (1989) 402.