

Combining metabolomic non-targeted GC×GC–ToF–MS analysis and chemometric ASCA-based study of variances to assess dietary influence on type 2 diabetes development in a mouse model

Saray Ly-Verdú · Thomas Maximilian Gröger · Jose Manuel Arteaga-Salas ·
Stefan Brandmaier · Melanie Kahle · Susanne Neschen · Martin Harbě de Angelis ·
Ralf Zimmermann

Received: 18 August 2014 / Revised: 25 September 2014 / Accepted: 30 September 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Insulin resistance (IR) lies at the origin of type 2 diabetes. It induces initial compensatory insulin secretion until insulin exhaustion and subsequent excessive levels of glucose (hyperglycemia). A high-calorie diet is a major risk factor contributing to the development of this metabolic disease. For this study, a time-course experiment was designed that consisted of two groups of mice. The aim of this design was to reproduce the dietary conditions that parallel the progress of IR over time. The first group was fed with a high-fatty-acid diet for several weeks and followed by 1 week of a low-fatty-acid intake, while the second group was fed with a low-fatty-acid diet during the entire experiment. The metabolomic fingerprint

of C3HeB/FeJ mice liver tissue extracts was determined by means of two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC–ToF–MS). This article addresses the application of ANOVA-simultaneous component analysis (ASCA) to the found metabolomic profile. By performing hyphenated high-throughput analytical techniques together with multivariate chemometric methodology on metabolomic analysis, it enables us to investigate the sources of variability in the data related to each experimental factor of the study design (defined as time, diet and individual). The contribution of the diet factor in the dissimilarities between the samples appeared to be predominant over the time factor contribution. Nevertheless, there is a significant contribution of the time–diet interaction factor. Thus, evaluating the influences of the factors separately, as it is done in classical statistical methods, may lead to inaccurate interpretation of the data, preventing achievement of consistent biological conclusions.

Published in the topical collection *Multidimensional Chromatography* with guest editors Torsten C. Schmidt, Oliver J. Schmitz, and Thorsten Teutenberg.

Electronic supplementary material The online version of this article (doi:10.1007/s00216-014-8227-4) contains supplementary material, which is available to authorized users.

S. Ly-Verdú · T. M. Gröger (✉) · J. M. Arteaga-Salas ·
R. Zimmermann

Joint Mass Spectrometry Centre, Cooperation Group
“Comprehensive Molecular Analytics”, Helmholtz Zentrum
Muenchen, 85764 Neuherberg, Germany
e-mail: thomas.groeger@helmholtz-muenchen.de

S. Ly-Verdú · T. M. Gröger · R. Zimmermann
Joint Mass Spectrometry Centre, Institute of Chemistry, Chair of
Analytical Chemistry, University of Rostock, 18057 Rostock,
Germany

S. Brandmaier
Institute of Epidemiology II, Helmholtz Zentrum Muenchen,
85764 Neuherberg, Germany

M. Kahle · S. Neschen · M. Harbě de Angelis
Institute of Experimental Genetics, Helmholtz Zentrum Muenchen,
85764 Neuherberg, Germany

Keywords Metabolomics · Chemometrics · Gas chromatography mass spectrometry · ANOVA-simultaneous component analysis (ASCA) · Type II diabetes · Mouse model

Introduction

In the complex hormone pathways of the body, a deficient insulin action leads, amongst others, to hyperglycemia (increased blood glucose levels), which is characteristic for the metabolic disease diabetes mellitus (DM). The pathogenic processes involved in the development of the most prevalent category of DM, i.e. type 2 DM (T2DM), arise from the combination of resistance to insulin action and an impaired compensatory insulin response [1]. Moreover, hepatic fat accumulation in non-alcoholic fatty liver disease (NAFLD)

is related to insulin resistance [2]. Amongst others, the factors contributing to the development of the insulin resistance and pancreatic β -cell dysfunction are high-calorie diet as well as sedentary lifestyle [3].

The *omics* technologies have been integrated into diabetes research as standard implement [4]. Metabolome—the entire set of metabolites within a biological system—is the downstream end product of the omics hierarchy and, therefore, an important source of information about protein expression manifested by gene regulation [5]. Metabolomics, the study of the metabolome, has demonstrated to be a powerful tool to investigate the biomarkers for disease diagnosis and risk prediction [6].

Multidimensional chromatography has proved to be a powerful and versatile tool for accurate metabolite separation. Furthermore, hyphenation with mass spectroscopy makes it a reliable and rapid system for tissue sample analysis, due to its capability for the analysis of complex mixtures and its utility for precise metabolite identification [7, 8]. In the present study, a two-dimensional gas chromatography time-of-flight mass spectrometry (GC $\times$ GC–ToF–MS) system was utilised for analysis. The superiority of GC $\times$ GC–ToF–MS over GC–MS lies in (1) its capacity to reduce coelution of chromatographic peaks—enhancing their structural elucidation and improving spectral purity—and (2) its capacity to reduce the limits of detection as well as to boost broader dynamic range [9].

The exploration of metabolomic data generated by multidimensional GC $\times$ GC–ToF–MS implicitly entails the difficulty of evaluating highly complex data. With this said, the versatility of chemometrics has proved to have an invaluable capacity to extract relevant information from this type of data [10].

For untargeted metabolomic studies, since each variable corresponds to a specific metabolite, the number of simultaneously measured inputs is of the order of thousands [11]. Using the statistical method of analysis of variance (ANOVA), which aim is to identify whether a significant difference exists amongst groups, it is possible to estimate the effect of the experimental factors on the data. However, for multivariate data, the use of multivariate ANOVA (MANOVA) is still not sufficient enough, as firstly, it does not explain the interrelation between variables, and secondly, it remains inadequate when the number of variables exceeds those of the measurements [12].

ANOVA-simultaneous component analysis (ASCA) procedure overcomes these handicaps, and it is gaining popularity in recent years, showing its usefulness as it combines the best of both ideas from ANOVA and principal component analysis (PCA) [13]. The principle of ASCA is to combine, on the one hand, the capacity to analyse separately different sources of experimental variation from ANOVA, and on the other hand, to explain the correlation amongst variables from PCA. An advantage of

ASCA over PCA is that the model is less affected by outliers, as the matrices are divided according to the different contributions to the variation in the data.

The aim of this study was to elucidate the influence of a high-fat dietary, at a metabolomic level, to a common mouse model (C3HeB/FeJ) and the contribution of the different experimental factors to the variances in the data. For this, efforts have been focused on the application of two-dimensional chromatographic techniques, such as GC $\times$ GC–ToF–MS, to detect early metabolic changes introduced by the diet time-course experiments. The findings of this study may lead to a contribution in the understanding of metabolic dysregulation in NAFLD. Likewise, ASCA has been applied as a new chemometric tool to increase the interpretability of the sources of variation in the dataset, which are an important issue when it comes to describe the understanding of the outcomes in biological experiments. To the best of our knowledge, no study to date encompasses the ASCA-based chemometric approach of two-dimensional chromatographic data together with the metabolomic assessment of the effect of diet in NAFLD progress in mouse model.

Experimental section

Chemical and reagents

All solvents were of analytical grade and purchased from Merck (Darmstadt, Germany) or Fluka Analytical Sigma-Aldrich (Steinheim, Switzerland). Water was obtained from a Millipore Milli-Q system (Billerica, MA, USA). Methoxamine hydrochloride was purchased from Supelco (Bellefonte, PA, USA) and *N*-methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA) from Macherey-Nagel GmbH (Düren, Germany). Stable isotope-labelled internal standard compounds, citric acid (2,2,4,4-D₄, 98 %), octanoic acid (D₁₅, 98 %) and fumaric acid (2,3-D₂, 98 %) were purchased from Cambridge Isotope Lab. Inc. (Andover, MA, USA); succinic acid (D₄, 98 %), pyruvic free acid (1-C₁₃, 95 %) and D-fructose (6,6-D₂, 98 %) were purchased from ISOTEC, Sigma-Aldrich (OH, USA). Ergosterol was purchased from Dr. Ehrenstorfer (Augsburg, Germany). The alkane standard mixture for the performance tests of GC systems C₈–C₂₀ was purchased from Fluka Analytical.

Animals and sample collection

All animals received humane care, and study protocols complied with the institution's guidelines. The Upper Bavarian district government (Regierung von Oberbayern, Gz.55.2-1-54-2532-4-11) approved all experimental procedures.

C3HeB/FeJ male mice (C3H; strain does not carry mouse mammary tumour virus or abnormal allele at Tlr4 locus but is

homozygous for retinal degradation allele Pde6brd1; The Jackson Laboratory, Maine, U.S.A.) were bred and housed under standard vivarium conditions (12:12 light–dark cycle). At an age of 14 weeks, male mice were single-housed, litter-matched and allocated into six groups. Three groups were switched to a high-fat diet (custom-made HFD, Ssniff, Soest, Germany) and three continued on a low-fat diet (LFD, Diet#1310, Altromin, Lage, Germany). Mice were anaesthetised and livers were then dissected, quickly freeze-clamped in liquid nitrogen and stored at -80°C until further processing.

As a safflower oil was the most prominent fat source for the high-fat diet (which contains n6-fatty acids), for the remainder of this paper, this diet will be designated as *safflower diet*, whilst the low-fat diet as *standard diet*.

A time-course experiment was designed in an attempt to recreate the conditions that may lead over time to the development of NAFLD, relating to insulin resistance.

Prior to the development of the illness, insulin resistance (IR) increases blood glucose levels in tissue, and, as the insulin response is inadequate, the patient will experience impaired glucose tolerance, which may progress to T2D. Even when hyperglycemia causes functional changes in tissues, it may not show clinical symptoms for a long period before DM detection. In prediabetic patients, hyperglycemia is developed relatively late in the pathogenesis of T2D. In fact, IR—and compensatory hypersecretion of insulin—begins well before glucose tolerance is impaired at all.

The intake plan consisted in 1, 2 and 3 weeks of a safflower oil diet, and 2 weeks of safflower oil diet followed by a reverse week of normal diet. Figure 1 abstracts the global design of this study.

Sample pretreatment

Snap-frozen livers were pulverised under liquid nitrogen. Samples of 25 mg were weighed into precooled 2-mL homogenisation tubes [14] containing 1.4-mm-diameter ceramic beads (Precellys, PeqLab, Germany) and immediately stored at -80°C and atmospheric pressure until extraction and analysis.

As internal standards, a mixture of 1 $\mu\text{g/mL}$ of [2,2,4,4- $^2\text{H}_4$]-citric acid, [$^2\text{H}_{15}$]-octanoic acid, [2,3- $^2\text{H}_2$]-fumaric acid, [$^2\text{H}_4$]-succinic acid, [1- ^{13}C]-pyruvic acid, [6,6- $^2\text{H}_2$]-D-fructose and ergosterol was added prior to homogenisation. Homogenisation was carried out in a Precellys 24 homogeniser (PeqLab Biotech. GmbH, Germany) equipped with an integrated cooling unit, three times operated at 5500 rpm ($\times g_{\text{av}}$) for 20 s with 30-s intervals, after addition of 1 mL of the solvent mixture (850 μL methanol, 100 μL ethanol and 50 μL water). A 2- μL aliquot of alkane standard mixture C:8–C:20 was also added in order to monitor the

chromatographic retention times. Samples were subsequently centrifuged in a Heraeus Multifuge 3SR (Thermo Scientific, Germany) at 10,000 rpm ($\times g_{\text{av}}$) for 10 min at -5°C . A total volume of 600 μL of supernatant liquid was then transferred to three 1-mL vials (200 μL each) and concentrated to complete dryness under a constant stream of nitrogen. Vials were then stored at -80°C until analysis.

The dried residues were resuspended in 50 μL of methoxamine hydrochloride (40 mg/mL pyridine) and incubated at 60°C for 60 min. After vortexing for 1 min, 50 μL of MSTFA was added and the samples were incubated at 60°C for 30 min. After a second vortex step, the derivatised samples were stored at room temperature for 180 min before injection [15].

Control samples consist of extract from 25 mg of a mixture of different livers from mice fed with normal laboratory diet on which the same pretreatment as mentioned for the samples was applied.

GC–MS analysis

For GC×GC–ToF–MS analysis, an Agilent 6890N gas chromatograph (Agilent Technologies, Palo Alto, CA, USA) equipped with a Leco GC×GC system (Leco Inc., St Joseph, MI, USA) and a dual-stage four-jet cryogenic (N_2) modulator (Leco Inc.) was used. The GC system was coupled to a Pegasus III time-of-flight mass spectrometer (Leco Inc.). A Combi PAL autosampler (CTC Analytics, Switzerland) performed 1- μL injections of derivatised extract followed by several 10- μL hexane and 10- μL dichloromethane wash cycles.

Controlled by the ATAS Evolution Workstation v.1.2a software, an OPTIC III injector (ATAS GL, Netherlands) was set up in splitless mode with an initial temperature of 70°C during injection. Starting 0.5 min after injection, the temperature was raised to 300°C at a rate of 2°C/s and subsequently held for 20 min. Initial injector head pressure set the head pressure to 380 kPa and maintained it during the 3.5-min transfer time. Subsequently, the column pressure was reduced to 170 kPa and then ramped parallel to the oven temperature program up to 320 kPa [9]. A programmed oven temperature started 3.5 min after injection, at a rate of 5°C/min , from 70 to 310°C and held until the end of the run. The GC×GC column set used for all the analysis was made up of a 1.5 m×250 μm deactivated fused silica tubing precolumn, combined with a 30 m×250 μm -internal-diameter (0.25 μm film thickness) 50 % phenylpolysilphenylene-siloxane (BPX-50, SGE, Australia) intermediately polar first-dimension column, coupled to a 1 m×180 μm (0.10 μm film thickness) trifluoro-propylmethyl polysiloxane (Rtx-200, Restek Corp., USA) selective second-dimension column. The transfer line was a 1 m×100 μm -internal-diameter deactivated fused silica tubing. A temperature offset of 100°C was set on the modulator with a 2-s modulation period.

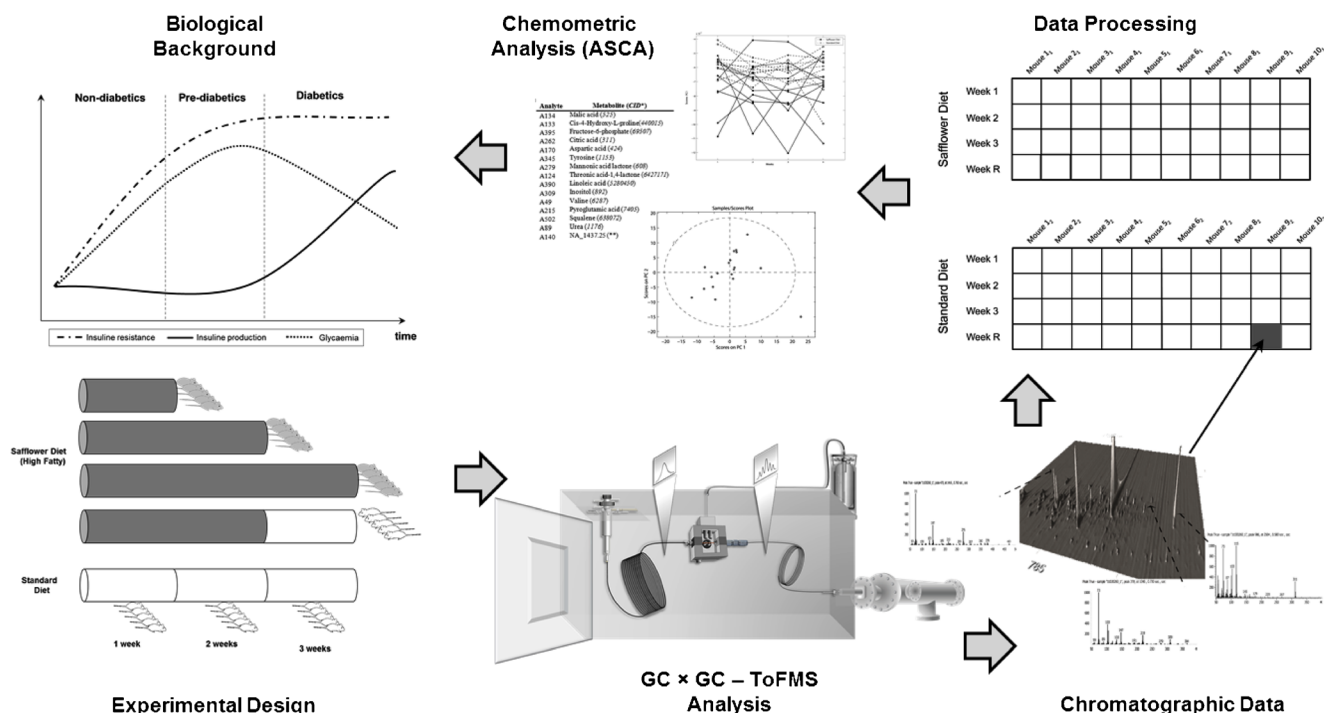


Fig. 1 Time-course design of the study. The experimental design recreates the typical IR progression over time sketched in the *upper graph*. Data from samples measured with GC×GC–ToF–MS were processed, and ASCA analysis was performed to discern the different sources of variation in them

The transfer line and ion source temperatures were kept at 270 and 250 °C, respectively, and the mass range was adjusted to 50–700 m/z at an acquisition rate of 100 spectra/s.

Alkane standard mixture was added to each replicate of the samples to monitor shifts in retention times and injection repeatability. Chromatographic performance was checked by means of Grob mixture injections [16]. Blank solvent injections were analysed to eliminate carryover effects through the sample analysis. To avoid systematic variations, the injection sequence was randomly set for analysis.

Data analysis

Data acquisition, analysis and processing were performed with ChromaTOF® software™, version 4.32 (Leco Inc.). Statistical analyses were performed in MATLAB® (R2011a, MathWorks, Inc., USA) using a home-written data preprocessing routine and the MATLAB-implemented PLS_Toolbox® (Eigenvector Research, Inc., USA). The parameters for spectrometric data processing were set with a signal-to-noise ratio of 50 and no smoothing. Deconvoluted and aligned peak data were exported via comma-separated values (.csv) format. The areas of matched peaks across the sample set were calculated and stored in a data matrix with columns representing the samples and rows representing the metabolites.

Data preprocessing

Deconvoluted peaks from the raw spectroscopic data were aligned by the Leco Software. To speed up and facilitate deconvolution and alignment, peaks corresponding to column bleed as well as determined derivatisation artefacts were excluded prior to the alignment.

Owing to the chosen threshold, wrong integration or qualitative differences in the metabolite profile, some values are stored as missing data. Missing values may lead to an over-estimation of the magnitude of the disparity in the data set. The presorting procedure for the aligned peaks included the identification of metabolites with missing values for more than half of the samples in every group, which were discarded from subsequent statistical analyses. For the remaining metabolites, missing values were replaced by the average of the non-missing values of every group separately. An exception was made for metabolites for which all values in one of the groups were zero. In those cases, the values were not replaced. The preprocessing revealed 401 features, which were used for further analysis. Peak areas were imported to MATLAB for further calculations.

Normalisation and instrumental optimisation

In order to monitor the repeatability of the chromatographic acquisition method, the relative standard deviations (RSD) for the retention times were calculated. For all aligned peaks,

including those that were later excluded, the first and the second retention time within all the samples showed an averaged RSD of 0.14 and 6.08 %, respectively.

Since deuterated internal standards were prior added to each sample, normalisation to their peak areas was applied to compensate for methodical and instrumental variations. First, the visual check of the chromatograms showed no endogenous interfering peaks as well as good separation of the standards and no matrix interfering peaks at their retention time. Evaluation of this procedure showed an averaged RSD of 10.25 %. Instrumental precision was determined by repeatability within day from 11 control samples, analysed every 24 h. By measuring the peak areas of each of the standards added, the averaged RSD was 8.5 %.

Multivariate analysis and statistics

Mean centering was applied to the data [17], also fulfilling the necessary and sufficient condition for the subsequent ASCA decomposition of the data matrix.

To perform PCA [18], data was leave-one-out cross-validated [19] and parameterised by means of the Pareto scaling method. Very high and low values of variation within variables calculated with PCA have an impact on the global variance. Thus, for the statistical comparison, normalised data was log-transformed to reduce the heteroscedasticity [20].

Disparities in the most relevant metabolites amongst samples evidence significant changes in the physiology of the mice. To discard those metabolites with only minor expressiveness, several filters have been applied to the data. First, the non-log-transformed loadings obtained for the first three principal components with a Euclidean distance less than 0.6 to the origin were not taken into consideration. Next, a density-based spatial clustering of applications with noise (DBSCAN) [21] (minPoint $s=3$, epsilon=0.06) was performed on the remaining metabolites, to exclude the scattered ones that cannot be assigned to a specific cluster.

A second statistical analysis was performed on the same data set. For this, normalised but not log-transformed data were structured in a matrix (Electronic Supplementary Material Fig S1) to enable the ASCA modelling [22].

In the linear model used in ANOVA that serves as the basis in ASCA, every measurement is decomposed into contributions to the variation caused by one treatment from variation caused by other sources, as described in Eq. (1).

$$x_{hki_j} = \mu_j + \alpha_{kj} + (\beta + (\alpha\beta))_{hkj} + (\alpha\beta\gamma)_{hki_j} \quad (1)$$

where, for every variable j , μ represents the mean value of all elements for the variable j , α_{kj} represents the contribution of the factor time, $(\beta + (\alpha\beta))_{hkj}$ represents the contribution of factor *treatment* and its interaction with *time*, $(\alpha\beta\gamma)_{hki_j}$

represents the variation specific to each *individual*. The indices are described in the Electronic Supplementary Material Fig. S1.

To allow the evaluation of the different experimental factors in complex experimental designs with multiple dependent variables, and as it is an overparameterised model, some constraints are imposed and listed in the Supplementary Table S1a.

Once the different factors have been isolated, the behaviour of the dependent variables under the different treatment levels is examined with PCA, which scores and loadings can be used to explain the decomposed data matrix, as described in Eq. (2).

$$\mathbf{X}_{hi_h} = \mathbf{1}\mathbf{m}^T + \mathbf{T}_K\mathbf{P}_1^T + \mathbf{T}_{Kh}\mathbf{P}_2^T + \mathbf{T}_{Khi_h}\mathbf{P}_3^T + \mathbf{E}_{hi_h} \quad (2)$$

where $\mathbf{1}$ is a $(K \times 1)$ vector of ones, $\mathbf{m}$ is a $(J \times 1)$ vector of the overall means of the peak areas, $\mathbf{T}$ are the scores of the principal components, $\mathbf{P}$ are the loadings of the principal components and $\mathbf{E}$ is the matrix of residuals. The matrix $\mathbf{X}_{hi_h}$ is separated into contributions from the overall mean ($\mathbf{1}\mathbf{m}^T$), the effect of the factor weeks ($\mathbf{T}_K\mathbf{P}_1^T$), the effect of interaction of diet with weeks ($\mathbf{T}_{Kh}\mathbf{P}_2^T$) and the effect of the interaction of diet, weeks and animal, i.e. contribution to the variation of each individual mouse, ($\mathbf{T}_{Khi_h}\mathbf{P}_3^T$).

By solving Eq. (3) under the given constraints listed in Supplementary Table S1b, which force the parts to be orthogonal to each other, the factors can be estimated separately.

$$\min_{\mathbf{m}, \mathbf{T}, \mathbf{P}} \sum_{h=1}^H \sum_{i_h=1}^{I_h} \|\mathbf{X}_{hi_h} - \mathbf{1}\mathbf{m}^T - \mathbf{T}_K\mathbf{P}_1^T - \mathbf{T}_{Kh}\mathbf{P}_2^T - \mathbf{T}_{Khi_h}\mathbf{P}_3^T\|^2 \quad (3)$$

For the remainder of this paper, time, treatment and individual factor will be designated as *week*, *diet* and *mouse* factor, respectively. Note that week represents the duration of the treatment (4 weeks) and not the ageing of the animals.

To verify how well the model fits the data, it is essential to check its goodness of fit. This procedure can be performed by means of QQ plots [23].

Results and discussion

PCA

Two principal components captured about 70 % of the variance within the data. Scores corresponding to the first and second principal components (PC) are depicted in Fig. 2a. On the first PC, which explains for about 50 % of the total variation, five separated clusters can be demarcated in the

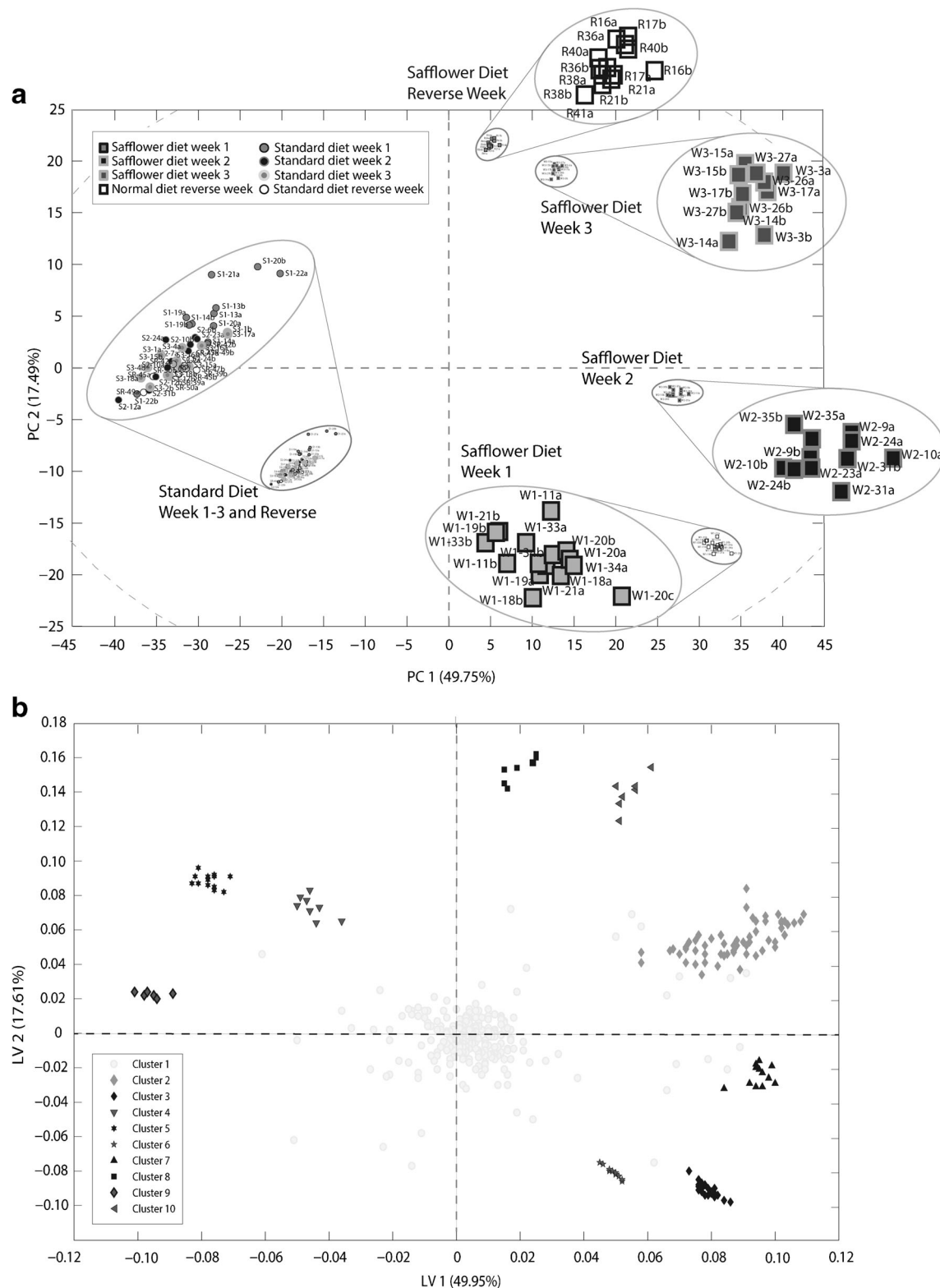


Fig. 2 **a** Distribution of the samples according to the PCA scores for the first two principal components. For each of the 4 weeks, *squares* represent high-fatty-acid diet (safflower) samples, whilst *circles* represent low-fatty-acid diet (standard) samples. The *dashed ellipse* encloses samples

with a Hotelling of 95 % confidence. **b** Loadings corresponding to the scores of PCA. Every *point* represents a specific metabolite. Numbers beside data points have been omitted for simplicity's sake

Table 1 Putative identification of metabolites corresponding to the loadings of the PCA procedure according to the cluster to which they belong

Analyte	Name (CID ^a)	Clusters
A1047	Phosphorylethanolamine (1015)	4
A155	Phosphoric acid (1004)	
A1364	all-Z-Docosahexaenoic acid (445580)	
A34	Ethylenethiourea (2723650)	
A88	NA_1123.4 ^b	6
A1247	Arachidonic acid (444899)	
A428	Proline (145742)	
A502	NA_1645.37	
A673	Lysine (5962)	2
A802	Erythrose (94176)	
A95	L-Valine (6287)	
A733	Glycerol (753)	
A623	NA_1776 ^b	7
A459	Pyroglutamic acid (7405)	
A789	Erythrose (94176)	
A977	Hexadecanoic acid (985)	
A455	Beta-D-glucopyranose (95629)	8
A740	Alpha-D-talofuranose (6431216)	
A1233	Eicosenoic acid (5282768)	
A838	Ascorbic acid (54670067)	
A39	Glycine (750)	9
A595	D-Glucose (79025)	
A1276	Linoleic acid (5280450)	
A674	Lysine (5962)	
A1614	Alpha-D-tochoperol (14985)	10
A239	Alpha-ionene (68057)	
A292	NA_1691.24 ^b	
A1077	Glucose-6-phosphate (5958)	
A610	NA_1831.58 ^b	3
A710	Dehydroascorbic acid (440667)	
A1608	Cholesterol (5997)	
A491	NA_1645.37 ^b	
A572	O-Phosphoethanolamine (1015)	Unknown
A601	D-Xylose (644160)	
A898	Unknown	

^a PubChem Compound Identifier^b Golm Metabolome Database

scatter plot. An obvious separation can be observed between samples on the left- and right-hand side from the zero dotted line for PC1. On the left bottom quadrant, samples corresponding to the standard diet are grouped together; thus, no discrimination is outstanding according to age of the mice in this group. This is a reasonable result, since these four groups of mice received the same intake of fatty acids during the entire experiment. The second PC, which explains for about 17 % of the total variation, presents a disparity for the samples corresponding to the first and second weeks of the safflower diet intake (right bottom quadrant) from those corresponding to the third and reverse weeks (right upper quadrant).

To investigate the contributions of the individual metabolites, we examined the particular loadings, underlying to this PCA. The loadings reflect the contribution and importance of a specific metabolite to the axes of the orthogonally transformed space. The referring loadings plot is shown in Fig. 2b. Each dot represents a specific metabolite, and its position reports its contribution to the principal component scores. Metabolites with a distance to the origin less than 0.6 (cluster 1) showed only minor contributions and are therefore assumed to reveal no significant changes in the metabolite profiles. All other metabolites assigned to several clusters (cluster 2 to cluster 10) are indicated with different markers and shadings. For those clusters, metabolites exhibiting significant contributions to the principal components are listed in Table 1.

ASCA modelling

ASCA divided the total variation into contributions corresponding to the experimental factors. Table 2 summarises the total percentage of explained variation for the model—first row. To display the result more legibly, for the following subsections, results from the submodels corresponding to every factor contribution, as well as the interaction between factors, are gathered in Table 2. The ranks for every submodel have been determined by means of a scree test [24], except for the time factor, where the rank was set to three as there is four measurement time points.

The percentage of each effect contributing to the sum of squares of the data matrix is listed in Table 1 (row 6). From

Table 2 Summary of variations obtained from the ASCA model and for each of its factor contribution

	Total percentage of explained variation	Percentage of explained variation for PC 1	Percentage of explained variation for PC2	Percentage of explained variation for PC3	Rank	Contribution percentage to the total data variation
Model	81.45	—	—	—	—	—
Submodel <i>K</i> (weeks)	97.70	52.28	30.68	17.04	3	22.7
Submodel <i>K_h</i> (diet)	100	76.07	15.32	—	2	44.3
Submodel <i>K_{h,i}</i> (animal)	67.01	23.14	27.97	—	2	14.0
Interaction (weeks–diet)		67.77	19.06	—	2	18.9

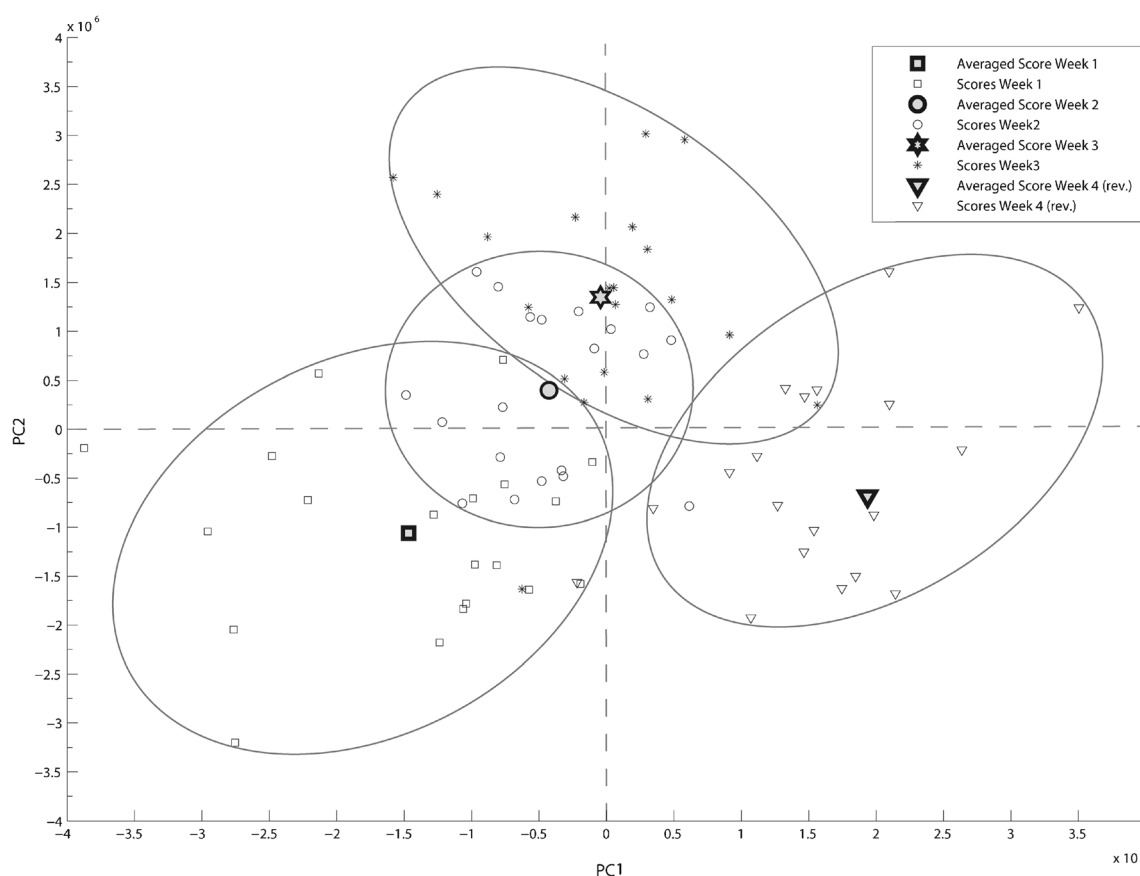


Fig. 3 ASCA model—submodel *K*: factor time scores. Scores on the first and second principal components and average scores for the 4 weeks

those values, it is clear that contribution from diet is the prevailing factor, although time has a non-negligible contribution to the total variation. It is noteworthy that the variation of interest for this experiment is related to weeks and to diet, and this is underlined by the percentage of contribution for the interaction of both to the total variation.

The Electronic Supplementary Material Fig. S7 shows the QQ plot of the residuals from the scores of the ASCA model and the residuals corresponding to the original variable, for four random chosen variables. Residuals from the ASCA model do not deflect from the dotted line, which represents a normal distribution, as the residuals from the original variable do. This indicates that the model is coherent with the theoretical assumptions.

ASCA modelling—factor week contribution (submodel *K*)

The submodel *K* captures all the variation that is only related to the duration of the experiment, since diet contribution is harvest in *diet* and *interaction* submodels.

The scores for the first and second components of the factor weeks are depicted in Fig. 3. For all mice, the average scores for each week (in bold) are well separated by the PC1. For the PC2, the first and reverse weeks show

Table 3 Putative identification of the metabolites corresponding to the ASCA procedure according to the factors for which they show a higher loading value. Factors t, d, i and t-d indicate time, diet, individual and interaction time–diet contribution, respectively

Number	Analyte	Name (CID ^a)	Factors			
21	A106	Glycerol (753)	t			
29	A1092	Linolic acid (5280450)			i	
62	A1248	Unknown				t-d
70	A129	Glycine (750)	t	d		t-d
120	A150	Unknown				t-d
132	A154	Phosphoric acid (1004)	t	d	i	t-d
143	A1608	Cholesterol (5997)	t	d	i	
163	A171	Urea (1176)		d		
212	A210	Unknown	t			
230	A292	Unknown	t			
231	A296	Unknown	t			
252	A355	5-Oxoproline (7405)		d	i	t-d
294	A585	D-Glucose (5793)	t	d		t-d
365	A88	NA_1123.4 ^b		d		t-d

^a PubChem Compound Identifier

^b Golm Metabolome Database

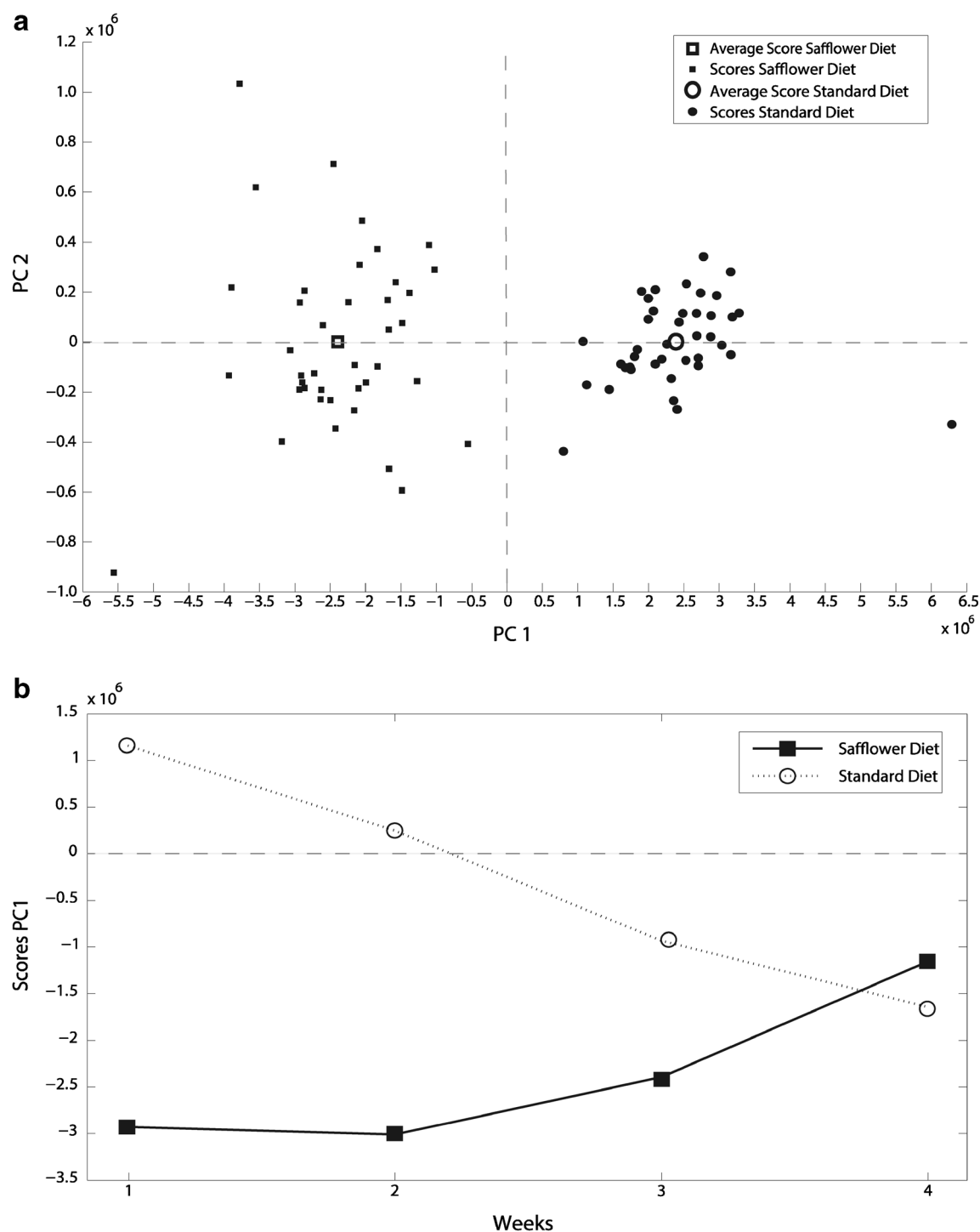


Fig. 4 ASCA model—submodel K_h : factor diet scores. **a** Scores on the first and second principal components and average scores for the two different diets. **b** Scores on the first principal component over the weeks for the two different diets

a similar score profile, as well as the second and third weeks. The second and third weeks present a change in the behaviour that is consistent with the expected development of the disease. Scores average in the reverse week of the normal diet could manifest a grade of recovery in concentration of certain metabolites responsible for the trend in previous weeks.

The distribution between the scores of the third component and the first and second components for this factor is given in the Electronic Supplementary Material Figs. S2 and S3, respectively. The loadings belonging to the first component are given in the Electronic Supplementary Material Fig. S4a. Table 3 lists the metabolites corresponding to the loadings for the first component.

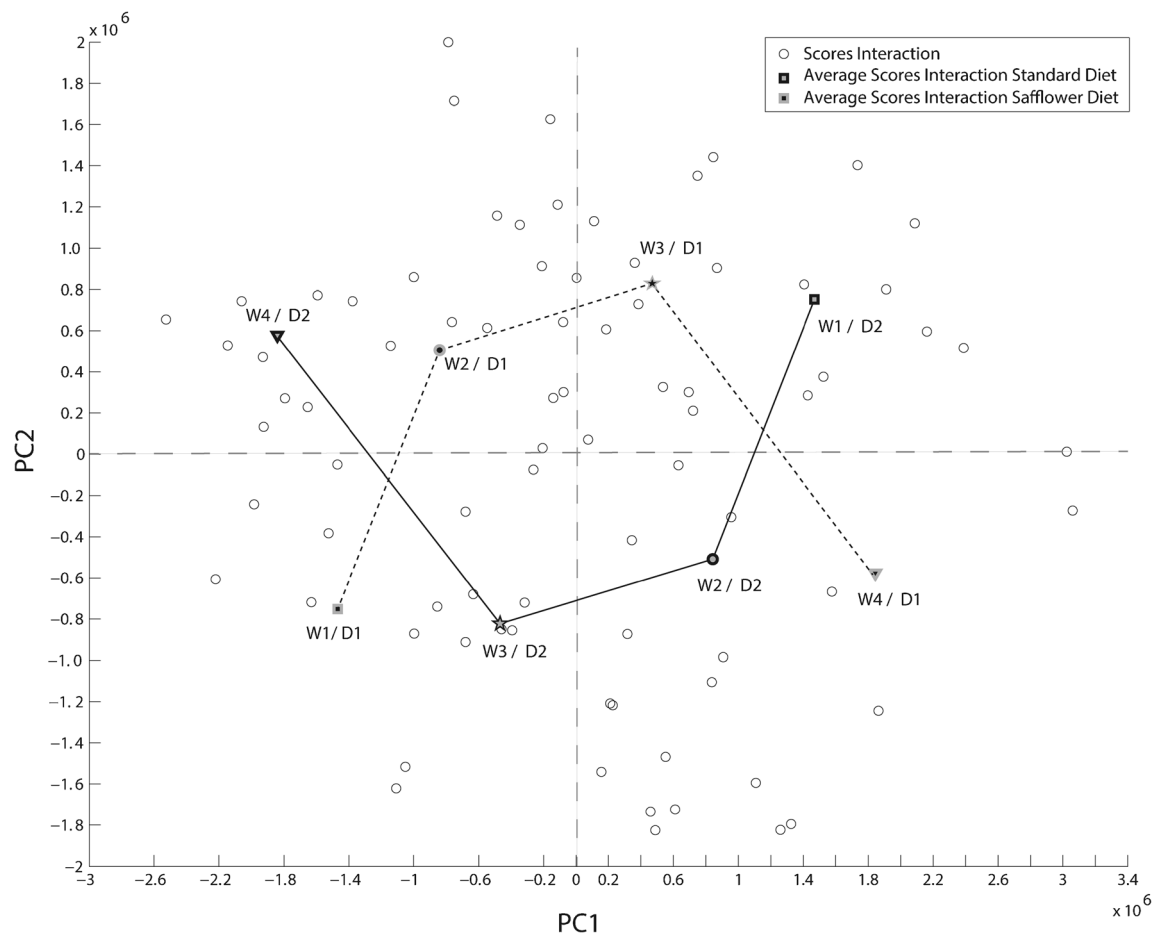


Fig. 5 ASCA model—submodel K_h : factor interaction time–diet scores on the first and second principal components and average scores for the ten mice of every week for the two different diets

As mice were already 14 weeks old at the feeding starting point, the actual age for them is 15, 16, 17 and 17 weeks for weeks 1, 2, 3 and 4, respectively. This implies that the time factor includes a “subfactor” that incorporates the previous ageing of the animal’s defendant to which group they belong. After 14 weeks, it is assumed that mice growing stage is completed and does not have an influence on the factor studied. This previous ageing has therefore not been taken into account in the model calculation; thus, the factor week is only representing the time during the dietary experiment.

The implementation of ASCA improves on PCA by ensuring that the variation captured by the submodel, namely week factor, is not originated from any other source but the studied factor. Despite that PCA identifies patterns highlighting their similarities, it does not distinguish between the factors that induce differences amongst samples.

ASCA modelling—factor diet contribution (submodel K_h)

The submodel K_h captures all the variation that is only induced by the different dietary, since week contribution is harvest in *week* and *interaction* submodels.

The scores for the first and second components, depicted in Fig. 4, represent the extent of the distinction of the safflower and standard diet groups from the factor weeks. Whilst the second principal component is not pertinent to distinguish one group from the other, for the first component, there is a clear separation between the two different diet groups. Furthermore, safflower diet scores are more disseminated than standard diet scores. Figure 4b shows scores on the first component for each of the weeks. Depending on which diet has been provided to the mice, an explicit pattern over the four time points is revealed. It is worth mentioning that for the reverse week, scores corresponding to both diet groups converge around a similar value. These results indicate that a high-fat diet should be considered as a factor by itself that affects the development of abnormalities in the studied liver composition. Moreover, a grade of recovery exists for the liver of mice that have been fed with normal laboratory diet after 14 days of fatty-acid feeding.

Again, comparing PCA with the ASCA modelling, the latter ensures that the variation explored in the submodel, mainly for the diet factor, is not influenced by the variation arising from the week factor.

The results are consistent with the conclusions presented by Kahle et al. [25]. In their study, hepatosteatosis (i.e. fatty liver) in C3H males is associated with hyperinsulinemia. The Electronic Supplementary Material Fig. S5 shows the ratio between high and low diet values for hepatic variables, during 21 days of experiments. Although the epididymal white adipose tissue (WAT) mass remains stable within the 3 weeks, liver triacylglycerides (TAG) show a gradual reduction from week to week, suggesting early adaptive responses in the liver. The progression of both the body mass (BM) and the body fat mass (BFM) follows a similar trend, with a drastic increase during the second week.

ASCA modelling—interaction week–diet contribution

The score plot for the interaction between the factor weeks and diet is depicted in Fig. 5. A crossover interaction between a line connecting the average scores of the safflower diet and the standard diet for the 4 weeks can be observed for the first as well as for the second component. This disordinal interaction indicates that the variable weeks has an impact on the different levels of the variable diet and vice versa. The more the scores differ in a week for the two different diets, the less significant is the mentioned interaction. Thus, the impact of time on the consequences of the diet, and conversely, becomes consecutively less significant in samples for the reverse week than for the first week, for the second week, and for the third week.

The loadings belonging to the first component for the factor diet and the interaction time–diet are given in the Electronic Supplementary Material Fig. S4b and S4d, respectively. Table 3 lists the metabolites corresponding to the loadings that induce this order.

ASCA modelling—factor mouse contribution (submodel $K_{h,i}$)

The standardisation of tissue sample from animals reminds a challenging task, as individuals may originate large variations in data. These particular fluctuations may lead to erroneous interpretation of the influence of other aimed factors, in our particular case, diet and week.

The scores for the first and second components, depicted in the Electronic Supplementary Material Fig. S6, explain the deviation of each mouse from the week–diet interaction. For the first component, the divergence within the profile of the mice is more conspicuous for the reverse week, primarily caused by samples corresponding to the standard diet (dotted line). However, for the second component, the disparity is only visible during the third week, primarily caused by samples corresponding to the safflower diet. Nevertheless, it has to be emphasised that the pattern for both groups is substantially steady, denoting no strong contribution of the factor mouse on the total variation.

The loadings belonging to the first component are given in the Electronic Supplementary Material Fig. S4c. Table 3 lists the metabolites corresponding to the loadings.

Conclusions

The influence of fatty diet over time on NAFLD development in mice liver has been investigated at a metabolomics level, by means of GC×GC–ToF–MS analysis. The application of ASCA chemometric methodology was proficient in elucidating the different factors contributing to the variation amongst samples, as well as their mutual interaction. The improvement in the interpretation of the sources of variation with ASCA over the PCA is manifest and has been illustrated for this series of real samples.

From the results, it can be deduced that diet has a major effect on the variation in data than weeks, even though the last factor plays an undeniable role in the global profile of the study. In fact, an interaction does exist between both factors that require consideration for a coherent interpretability of biological outcomes. Moreover, metabolite concentration underlying the biological response suggests certain reversibility in the progress of the disease after reintroducing a non-fatty diet for 1 week. The model also ratifies the assumption that individual contribution from each animal is not the main cause of differences found between profiles over the weeks.

References

1. Stumvoll M, Goldstein BJ, van Haeften TW (2005) Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* 365(9467):1333–1346. doi:10.1016/S0140-6736(05)61032-X
2. Marchesini G, Brizi M, Morselli-Labate AM, Bianchi G, Bugianesi E, McCullough AJ, Forlani G, Melchionda N (1999) Association of nonalcoholic fatty liver disease with insulin resistance. *Am J Med* 107(5):450–455. doi:10.1016/S0002-9343(99)00271-5
3. DeFronzo RA (2009) From the triumvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. *Diabetes* 58(4):773–795. doi:10.2337/db09-9028
4. Bain JR, Stevens RD, Wenner BR, Ilkayeva O, Muoio DM, Newgard CB (2009) Metabolomics applied to diabetes research: moving from information to knowledge. *Diabetes* 58(11):2429–2443. doi:10.2337/db09-0580
5. Patti GJ, Yanes O, Siuzdak G (2012) Innovation: metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* 13(4):263–269
6. Xia J, Broadhurst D, Wilson M, Wishart D (2013) Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* 9(2):280–299. doi:10.1007/s11306-012-0482-9
7. Almstetter M, Oefner P, Dettmer K (2012) Comprehensive two-dimensional gas chromatography in metabolomics. *Anal Bioanal Chem* 402(6):1993–2013. doi:10.1007/s00216-011-5630-y

8. Wu Z, Huang Z, Lehmann R, Zhao C, Xu G (2009) The application of chromatography-mass spectrometry: methods to metabonomics. *Chromatographia* 69:23–32. doi:[10.1365/s10337-009-0956-8](https://doi.org/10.1365/s10337-009-0956-8)
9. Koek M, van der Kloet F, Kleemann R, Kooistra T, Verheij E, Hankemeier T (2010) Semi-automated non-target processing in GC×GC–MS metabolomics analysis: applicability for biomedical studies. *Metabolomics* 7(1):1–14. doi:[10.1007/s11306-010-0219-6](https://doi.org/10.1007/s11306-010-0219-6)
10. Zeng Z-D, Hugel H, Marriott P (2011) Chemometrics in comprehensive multidimensional separations. *Anal Bioanal Chem* 401(8):2373–2386. doi:[10.1007/s00216-011-5139-4](https://doi.org/10.1007/s00216-011-5139-4)
11. Boccard J, Rudaz S (2014) Harnessing the complexity of metabolomic data with chemometrics. *J Chemom* 28(1):1–9. doi:[10.1002/cem.2567](https://doi.org/10.1002/cem.2567)
12. Smilde AK, Jansen JJ, Hoefsloot HCJ, Lamers R-JAN, van der Greef J, Timmerman ME (2005) ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics* 21(13):3043–3048. doi:[10.1093/bioinformatics/bti476](https://doi.org/10.1093/bioinformatics/bti476)
13. Zwanenburg G, Hoefsloot HCJ, Westerhuis JA, Jansen JJ, Smilde AK (2011) ANOVA–principal component analysis and ANOVA–simultaneous component analysis: a comparison. *J Chemom* 25(10):561–567. doi:[10.1002/cem.1400](https://doi.org/10.1002/cem.1400)
14. Römisch-Margl W, Prehn C, Bogumil R, Röhring C, Suhre K, Adamski J (2011) Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* 8:1–10. doi:[10.1007/s11306-011-0293-4](https://doi.org/10.1007/s11306-011-0293-4)
15. Birkemeyer C, Kolasa A, Kopka J (2003) Comprehensive chemical derivatization for gas chromatography-mass spectrometry-based multi-targeted profiling of the major phytohormones. *J Chromatogr A* 993(1–2):89–102. doi:[10.1016/S0021-9673\(03\)00356-X](https://doi.org/10.1016/S0021-9673(03)00356-X)
16. Grob K, Grob G, Grob K Jr (1981) Testing capillary gas chromatographic columns. *J Chromatogr A* 219(1):13–20
17. Bro R (2003) Multivariate calibration: what is in chemometrics for the analytical chemist? *Anal Chim Acta* 500(1–2):185–194. doi:[10.1016/S0003-2670\(03\)00681-0](https://doi.org/10.1016/S0003-2670(03)00681-0)
18. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemom Intell Lab Syst* 2(1–3):37–52. doi:[10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
19. Trygg J, Holmes E, Lundstedt TR (2006) Chemometrics in metabonomics. *J Proteome Res* 6(2):469–479. doi:[10.1021/pr060594q](https://doi.org/10.1021/pr060594q)
20. van den Berg R, Hoefsloot H, Westerhuis J, Smilde A, van der Werf M (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 7(1):142. doi:[10.1186/1471-2164-7-142](https://doi.org/10.1186/1471-2164-7-142)
21. Ester M, Kriegel H, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Simoudis E, Han J, Fayyad U (eds) Second International Conference on Knowledge Discovery and Data Mining. AAAI Press, pp 226–231. doi:[citeulike-article-id:2265233](https://doi.org/10.1145/234253.234273)
22. Jansen JJ, Hoefsloot HCJ, van der Greef J, Timmerman ME, Westerhuis JA, Smilde AK (2005) ASCA: analysis of multivariate data obtained from an experimental design. *J Chemom* 19(9):469–481. doi:[10.1002/cem.952](https://doi.org/10.1002/cem.952)
23. Stehlík M, Střelec L, Thulin M (2014) On robust testing for normality in chemometrics. *Chemom Intell Lab Syst* 130(0):98–108. doi:[10.1016/j.chemolab.2013.10.010](https://doi.org/10.1016/j.chemolab.2013.10.010)
24. Cattell RB (1966) The scree test for the number of factors. *Multivar Behav Res* 1(2):245–276. doi:[10.1207/s15327906mbr0102_10](https://doi.org/10.1207/s15327906mbr0102_10)
25. Kahle M, Horsch M, Fridrich B, Seelig A, Schultheiß J, Leonhardt J, Irmeler M, Beckers J, Rathkolb B, Wolf E, Franke N, Gailus-Durner V, Fuchs H, de Angelis MH, Neschen S (2013) Phenotypic comparison of common mouse strains developing high-fat diet-induced hepatosteatosis. *Mol Metab* 2(4):435–446. doi:[10.1016/j.molmet.2013.07.009](https://doi.org/10.1016/j.molmet.2013.07.009)