A laboratory approach for characterizing chronic fatigue: what does metabolomics tell us?

Elardus Erasmus[1,*], Shayne Mason[1], Mari van Reenen[1], Francois E. Steffens[2], B. Chris Vorster[1] and Carolus J. Reinecke[1]

1 Centre for Human Metabolomics, Faculty of Natural and Agricultural Sciences, North-West University (Potchefstroom Campus), Private Bag X6001, Potchefstroom, South Africa
2 Department of Consumer Science, University of Pretoria, Pretoria, South Africa

*Correspondence author: Elardus Erasmus. lardus.erasmus@nwu.ac.za

Other email addresses:
Shayne Mason. nmr.nwu@gmail.com
Mari van Reenen. 12791733@nwu.ac.za
Francois E. Steffens. steffensfe@gmail.com
B. Chris Vorster. chris.vorster@nwu.ac.za

Abstract

Introduction: Manifestations of fatigue range from chronic fatigue up to a severe syndrome and myalgic encephalomyelitis. Fatigue grossly affects the functional status and quality of life of affected individuals, prompting the World Health Organization to recognize it as a chronic non-communicable condition.

Objectives: Here, we explore the potential of urinary metabolite information to complement clinical criteria of fatigue, providing an avenue towards an objective measure of fatigue in patients presenting with the full spectrum of fatigue levels.

Methods: The experimental group consisted of 578 chronic fatigue female patients. The measurement design was composed of (1) existing clinical fatigue scales, (2) a hepatic detoxification challenge test, and (3) untargeted proton nuclear magnetic resonance (^{1}H-NMR) procedure to generate metabolomics data. Data analysed via an in-house Matlab script that combines functions from a Statistics and a PLS Toolbox.

Results: Multivariate analysis of the original 459 profiled ^{1}H-NMR bins for the low (control) and high (patient) fatigue groups indicated complete separation following the detoxification experimental challenge. Important bins identified from the ^{1}H-NMR spectra provided quantitative metabolite information on the detoxification challenge for the fatigue groups.

Conclusions: Untargeted ^{1}H-NMR metabolomics proved its applicability as a global profiling tool to reveal the impact of toxicological interventions in chronic fatigue patients. No clear potential biomarker emerged from this study, but the quantitative profile of the phase II biotransformation products provide a practical visible effect directing to up-regulation of crucial phase II enzyme systems in the high fatigue group in response to a high xenobiotic-load.

Keywords: Chronic fatigue; Detoxification challenge test; Piper fatigue scale; ^{1}H-NMR metabolomics; Phase II biotransformation

Introduction

Fatigue is one of the most common conditions reported by patients in primary care practice. Fatigue frequently co-occurs with pain, sleep disturbances and depression prompting medical practitioners to use diverse definitions and diagnostic labels to describe possible underlying causes of these manifestations. Thus, fatigue appears as a symptom of idiopathic conditions, at the one end of the fatigue spectrum, towards chronic fatigue syndrome (CFS) and myalgic encephalomyelitis (ME), at the other end. In its more persistent form, fatigue is now recognized by the World Health Organization (WHO) as one of the chronic non-communicable conditions, which are rapidly becoming endemic worldwide, recognizing that fatigue is associated with a significant decline in the functional status and quality of life (QOL) of affected individuals.

Although a diagnosis is essential for providing appropriate care, there is no established diagnostic test for any of the manifestations of fatigue. To this end, various wellness and health institutions in the United States (US) tasked the Institute of Medicine (IOM) to develop evidence-based clinical diagnostic criteria at least for ME/CFS and to recommend whether new terminology for ME/CFS should be adopted (Clayton 2015). Following extensive research involving patients and clinicians, the IOM committee proposed redefined diagnostic criteria to facilitate timely diagnosis and to improve understanding of the illness among healthcare professionals and the public (Clayton 2015). The diagnostic criteria indicative of CFS/ME specify: (1) a substantial reduction or impairment in the ability to engage in pre-illness levels of activity (occupational, educational, social or personal life); (2) symptoms that present for more than 6 months; (3) episodes of profound fatigue; (4) symptoms of new onset and not resulting from on-going or unusual excessive exertion; and (5) symptoms that are nor substantially alleviated by rest. From this descriptive, it is clear that clinical criteria, though relatively subjective, remain key in the diagnosis of chronic fatigue seen in CFS/ME and that no objective measure for fatigue in general has been proposed—the issue that is addressed in this paper.

Seeing the multifactorial nature of chronic fatigue, it seems highly unlikely that an objective measure, as a verified a single biomarker or biosignature comprised of a limited number of biomarkers, will be found for fatigue conditions, unless the essence of this disorder, its causes and pathophysiology becomes clearly delineated. However, it has recently been shown (Erasmus et al. 2019) that data from a battery of conventional tests provided some objective indicators to complement clinical and lifestyle data which enabled the classification of a cohort of patients into eleven subgroups, ranging from low to high fatigue. In addition, metabolomics research has provided insights into disorders of comparable complexity to chronic fatigue, like fibromyalgia syndrome (Hackshaw et al. 2018; Malatji et al. 2017), chronic widespread pain (Freidin et al. 2018; Hadrévi et al. 2015) and irritable bowel syndrome (Fourie et al. 2016; Ponnusamy et al. 2011). Likewise, metabolomics approaches using intervention studies proved to provide valuable insights into experimental studies on nutrition (Wittwer et al. 2011), detoxification (Irwin et al. 2016) and acute alcohol consumption (Irwin et al. 2018). Here we propose a laboratory approach to facilitate an objective evaluation of fatigue in a clinically selected group of chronic fatigue patients. The approach is compiled of: (1) existing clinical scales to score (Piper et al. 1998); (2) a known hepatic detoxification challenge test (Cordts et al. 2001) and (3) untargeted

proton nuclear magnetic resonance (^{1}H-NMR) procedure to generate metabolomics data.

The selection of cases used for the laboratory assessment was made from 576 females identified by clinicians as suffering from chronic fatigue, supported by information from two known assessment scales. Based on exclusion criteria, the control (low fatigue) and patient (high fatigue) groups were selected from this cohort for the metabolomics study. Fatigue in our patient group was scored based on the Piper Fatigue Scale (PFS) (Piper et al. 1998) developed for fatigue prevailing in oncology patients. The information from the PFS was supplemented with that from a general Medical Symptoms Questionnaire (MSQ) used by the Departments of Medicine, Mercy Hospital and Maine Medical Center, Portland, for comprehensive profiling of patients with idiopathic conditions.

The hepatic detoxification test probes the gut-liver function through a hepatic detoxification challenge with acetaminophen and acetylsalicylic acid (Cordts et al. 2001). The laboratory instrument used for the objective measurement of fatigue is based on the biotransformation profile derived from these hepatic challenge tests. The underlying physiological assumption is that fatigue is a symptom of energy depletion (e.g. as indicated by the PFS). Based on the presumed causal relationship between exogenous stimuli, biotransformation responses and fatigue, we speculated that the response to the highly energy dependent hepatic challenge test (that is, ATP required for the in vivo synthesis of biotransformation products) should be more pronounced in controls than in fatigue patients supposed to suffer from energy depletion.

In this exploratory study we followed a holistic approach to determine the response to an intervention by using metabolomics technology in the generation and analysis of untargeted ^{1}H-NMR spectral data, produced prior to and following the intervention. Important bins identified from the ^{1}H-NMR spectra provided qualitative information to compare the effect of the challenge tests on control and fatigue cases and used to identify and quantify biotransformation markers associated with fatigue.

The IOM's report on fatigue emphasized the need for increased focus on the complex phenomenon of fatigue, paving the way to learn more on conditions covering the full spectrum of fatigue conditions and to diagnose and treat patients suffering from a disease that severely affects their QOL. The outcomes of the present metabolomics study suggest that the intervention used can provide an objective measure to distinguished the response of patients with low and high fatigue, and has the potential, given further refinement, to benefit patients suffering from fatigue conditions.

Materials and methods

Objective, ethical approval and case selection

The primary objective of the study was to propose a laboratory-based procedure that could pave the way towards an objective indicator of fatigue in patients observed as such by physicians in general clinical practice. The point of departure for the design of the study was thus to enrol a large number of patients suffering from chronic fatigue. Next, all prospective participants were informed of the objective of the study, which complied with

the requirements of the appropriate Ethics Committee of North-West University (Reference no. NWU-00102-12-A1). All participants provided written consent (see Supplementary Information (SI) for Informed consent forms) to participate and for eventual publication of anonymized results in accordance with the ethical requirements of the university.

Patients suffering from fatigue emanating from well-defined clinical conditions such as cancer and CFS/ME or other known conditions such as diabetes or hypertension were excluded from the study. Exclusions were based on the assessment of the physician and the outcome of the MSQ (see SI on Patient procedures and selection for full details). Next, participants were asked to complete a self-report questionnaire that included a demographic profile and an English version of the revised Piper Fatigue Scale (PFS). The PFS used consisted of 22 numerical items that assess fatigue experienced by the patients. All items were coded on a 0–10 numeric scale and resolve into four dimensions of subjective fatigue: behavior/severity, affective meaning, sensory, and cognitive/mood. In total, 673 women with complaints of fatigue were assessed and of these 576 were found to be eligible for the study. The age, height, weight or BMI did not have a confounding effect on the biochemical data or on the classification of fatigue groups, indicated by a correlation analysis to indicate potential confounding effects (Erasmus et al. 2019).

Having applied the PFS to a clinically different fatigue group than the instrument was designed for, required a re-evaluation of the fatigue dimension. To this end, various combinations of factor extraction methods (principal components and principal axis factoring) and rotation methods (varimax rotation and direct oblimin rotation) were employed. The different approaches all indicated two underlying factors in our data set. The first factor was seen to load mainly on items 2–17 and was termed Energy Fatigue, whereas the second factor loaded mainly on items 18–23 and was termed Mental Fatigue. The factor scores were calculated as based on average scores with Energy Fatigue score the mean of items 2–17 (16 items) and Mental Fatigue score the mean of items 18–23 (6 items). The reliability of the measurement for the two factors was evaluated by means of Cronbach's Alpha and proved to be high (0.974 and 0.971 for the 16 and 6 energy and mental fatigue factors, respectively).

The MSQ contains 19 domains, most with four items, totalling 75 indicators to each of which the patient assigns a Likert scale numerical ranking from 0 ("never" or "seldom") to 4 ("frequently with severe effect"). Note that one question, found in the energy section about hyperactivity, was omitted to enhance the reliability (see SI for further details). For this data set, factor analysis rotation and extraction methods did not converge, while PCA did not provide useful outputs. It was therefore decided to calculate factor scores based on the sections and total since the MSQ was applied in a similar context to that in which it was developed. The MSQ domain scores were calculated as the mean of the items per domain, in addition the items were combined into an overall MSQ score.

Two groups were selected for the hepatic challenge tests: (1) A low fatigue group consisted of 39 cases (control group), scoring lower than 4.0 on energy as well as mental fatigue factors and an overall score from the MSQ not exceeding 28. (2) A high fatigue group also

comprised 31 subjects (patient group), scoring above 7.0 on energy as well as mental fatigue factors and an overall score from the MSQ exceeding 147. The distribution of these patients within their fatigue categories (low and high) is shown in Fig. 1.

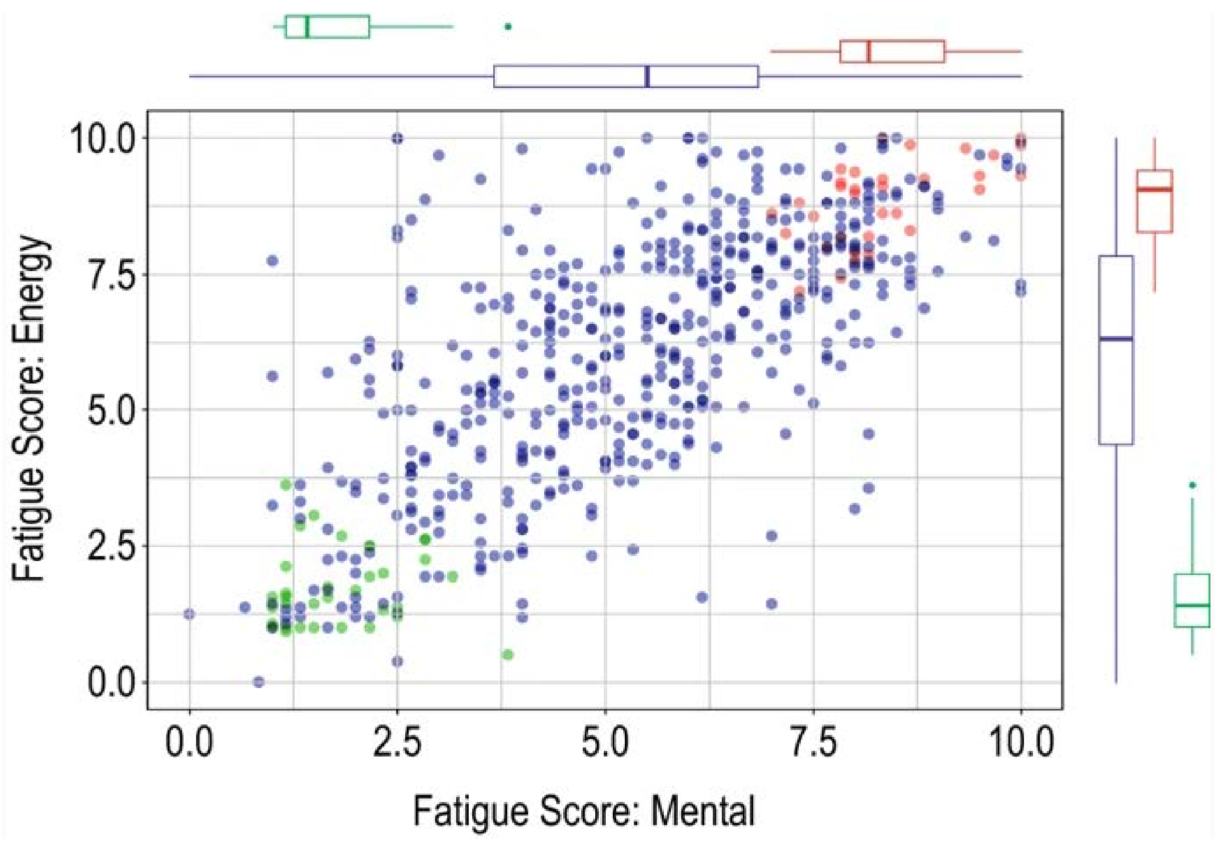

Fig. 1. Distribution of all selected cases based on the individual's scores on energy fatigue and mental fatigue factors derived from the PFS. The scores are categorized as low level fatigue (0 < score < 4), medium level of fatigue (4 ≤ score < 7) and high level of fatigue (score ≥ 7). The distribution of all 576 cases is shown as blue, green or red dots. The green dots indicate the group of 39 cases with low fatigue (PFS energy and mental < 4.0 and overall MSQ ≤ 28). The red dots indicate the high fatigue group of also 31 subjects (PFS energy and mental > 7.0 and overall MSQ > 147). Box plots indicate the distribution of scores which each group as evident from the corresponding colours

Laboratory methods: The challenge tests and metabolomics data generation

The two hepatic challenge tests were performed simultaneously. Prior to the challenge, a base urine sample was collected from each individual, and stored as − 20 °C till analyses (pre-sample). The first challenge involved two aspirin (600 mg) and two acetaminophen (paracetamol) tablets (1000 mg) taken at 21h00 in the evening and overnight and early morning (ending at 7h00) urine samples collected in a special container provided in the test kit and stored at − 20 °C as well (post-sample).

^{1}H-NMR analysis

A volume of 600 µL of urine was centrifuged at 12,000*g* for 5 min. Of the supernatant, 540 µL was collected in a microcentrifuge tube, with 60 µL NMR buffer solution [1.5 M potassium phosphate solution in deuterium oxide with internal standard TSP (trimethylsilyl-

2,2,3,3-tetradeuteropropionic acid); pH 7.4]. Sample was mixed under vortex to ensure completely homogenous and centrifuged again at 12,000*g* for 5 min. A volume of 540 µL of supernatant was transferred to a 5 mm NMR tube for analysis.

Samples were measured, randomized, at 500 MHz on a Bruker Avance III HD NMR spectrometer equipped with a triple-resonance inverse (TXI) 1H{15N, 13C} probe head and x, y, z gradient coils. ^{1}H spectra were acquired as 128 transients in 32 K data points with a spectral width of 12,000 Hz. The sample temperature was maintained at 300 K and the H_2O resonance was presaturated by single-frequency irradiation during a relaxation delay of 4 s, with a 90° excitation pulse of 8 µs. Shimming of the sample was performed automatically on the deuterium signal. The resonance line widths for TSP and metabolites were < 1 Hz. Fourier transformation and phase and baseline correction were done automatically. Software used for NMR processing was Bruker Topspin (V3.5). Bruker AMIX (V3.9.14) was used for metabolite identification and quantification. Cases were quantified across 459 equally spaced bins.

Statistical analysis

To aid interpretation two-group comparisons were performed within high and low fatigue groups as pre vs post intervention pairs. The data were pre-processed in subsets associated with each two-group comparison. First, bins with 50% or more zero-valued observation in both groups were removed. Second, the remaining zero-valued observations were replaced by random numbers generated uniformly below the lowest non-zero observation. Finally, the data were scaled by first performing a log transformation to adjust for asymmetry in distribution, since metabolomics data are known to present with a positively skewed distribution. Following this transformation, the data were auto-scaled to adjust for the effect of differences in abundance, since bins in high abundance are not necessarily of greater importance.

Univariate and multivariate methods were applied to allow for a comprehensive overview of results. The univariate method used to assess the statistical significance of differences between the dependent groups was the non-parametric Wilcoxon test. Effect sizes were used to assess the magnitude of significant differences or practical significance. Effect sizes were calculated as the respective test statistic scaled to the sample size. Scatter plots colour coded to show significance, often referred to as volcano plots, were used to summarize statistical and mean difference ratios or fold changes across bins.

Unsupervised multivariate methods were used in support of supervised methods as a form of validation. Unsupervised PCA and cluster analysis were performed to project the data onto fewer more manageable dimensions and to show purely data driven associations between cases, respectively. Euclidean distance with Ward linkage was used to form hierarchical clusters. PCA scores plots were used to explore whether a large proportion of variation in the data can be associated with the group structure by overlaying confidence intervals for groups. These methods serve to support or caution against the interpretation of similar scores plots derived using partial least squares discriminant analysis (PLS-DA), a supervised method used to emphasize any differences between groups. Again, confidence intervals were used to assess discriminatory ability.

All pre-processing and statistical analysis were performed using an in-house Matlab script that serves to combine functions from the Statistics Toolbox [MATLAB with Statistics and Toolbox Release 2016b, The MathWorks, Inc., Natick, MA, USA)] and the PLS Toolbox as provided by EigenVector [PLS-Toolbox 8.2.1 (2016). Eigenvector Research. Software available at www.eigenvector.com.].

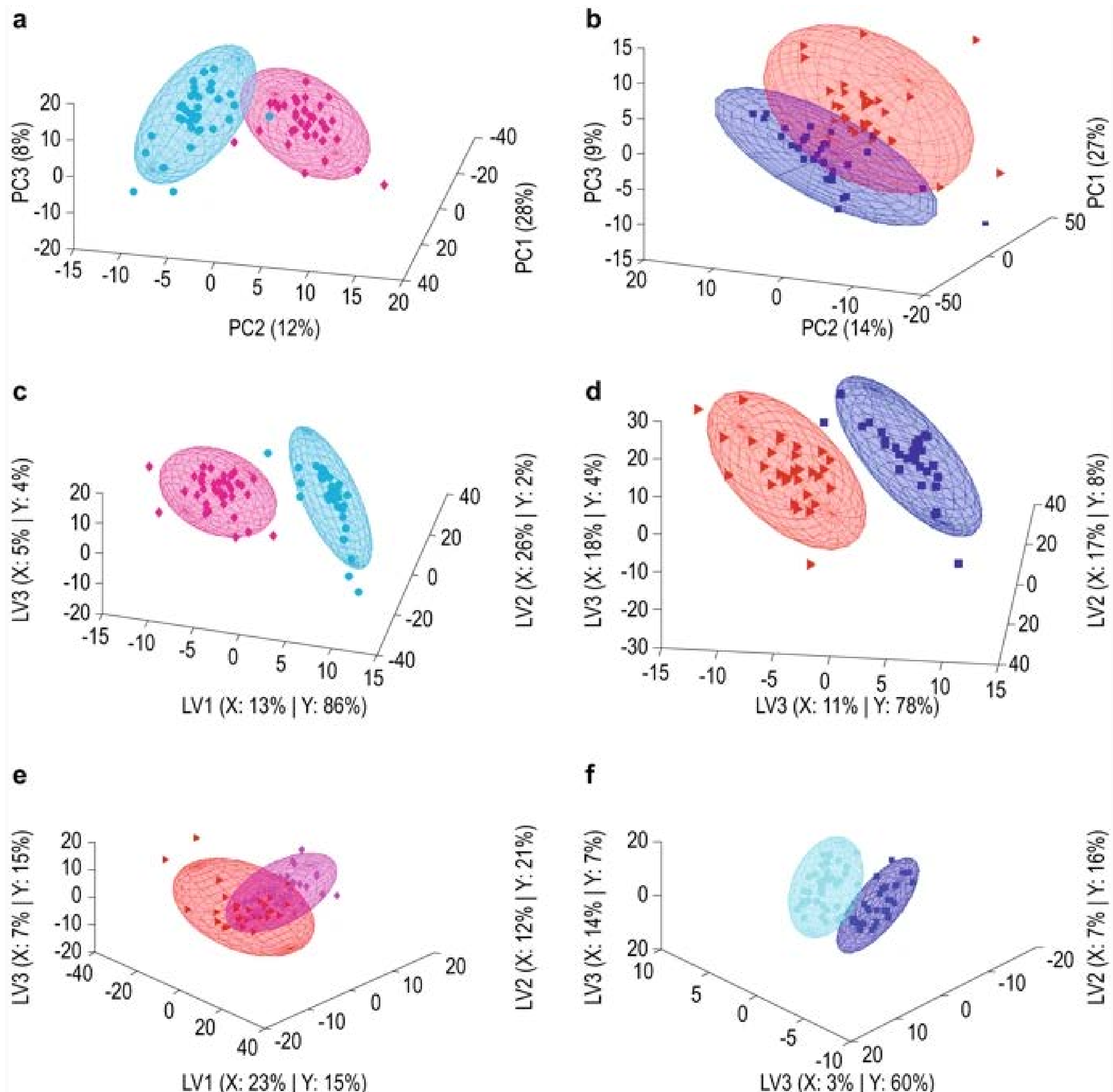

Fig. 2. Group separation between experimental groups through multivariate analysis based on equidistant binning data. PCA indicating a clear change between pre and post intervention profiles for low (**a**) and high (**b**) fatigue cases, similarly for PLS-DA models for low (**c**) and high (**d**) fatigue cases. **e** The PLS-DA models for the low and high fatigue groups for the post intervention data. **f** The PLS-DA models from each participant's relative change before (pre) and after (post) the intervention, shown for the low and high fatigue groups. Coloured areas represent 90% confidence intervals (CI) for groups

Results

Qualitative metabolomics information following the hepatic challenges

The qualitative data was obtained from the original 459 ^{1}H-NMR profiled bins for the experimental groups before and after the intervention. Supposed changes in metabolite profiles from the two patients groups following the interventions were firstly established through two multivariate approaches. For the bin data, the group separations before and after the challenge were investigated using unsupervised PCA (Fig. 2a and b), and supervised PLS-DA (Fig. 2c and d). The PCA showed a virtual clear separation for the low fatigue group prior to and after the intervention. For the high fatigue group only a partial separation was detected. As expected, complete separation for both the low and high fatigue groups were detected through the PLS-DA. As this method of classification tends to overfits the data (Westerhuis et al. 2008), the PCA serves as some verification, but taken with the predictive accuracy $R^2(Y)$ of the PLS-DA models and its leave-one-out cross-validated counterparts $Q^2(Y)$, some surety can be attainted. The $R^2(Y)$ and $Q^2(Y)$ values were 90.1% and 79.8% for the high fatigue group model and 93.1% and 86.3% for the low, respectively. Interestingly, both the PCA and PLS-DA models appear to have a reduced ability to discriminate pre and post intervention measures in the high fatigue group, which may support the idea that fatigue hinders the ability to respond to a metabolic challenge. Similar findings were noted in the univariate analysis below.

An incomplete group separation in the PLS-DA models was observed between low and high fatigue subjects following the hepatic detoxification test, as shown in Fig. 2e. This observation suggests subtle differences in metabolite concentrations between low and high fatigue groups, following the intervention, which will be further investigated through a quantitative comparison of the biotransformation products, as will be shown below. Figure 2f indicates the relative change before (pre) and after (post) exposure to the intervention stressors for the high and low fatigue groups. Relative change was computed in a similar manner to a fold change but based on the two observations of each participant. The relative change values were compared across fatigue groups using a PLS-DA model. The resulting scores plots indicate a difference in response of the individuals representing the low and high fatigue groups.

Next, univariate analyses using Wilcoxon *p*-values and fold changes on the same bin data were determined and summarized in the volcano plots for the low (a) and high (b) fatigue groups (Fig. 3). The outcome of these analyses of the bin data indicate large-magnitude changes (fold change: $|\log_2 FC| > 1.5$) that are also statistically significant (Wilcoxon test: $p \leq 0.05$), directing to several metabolites of importance that may cause the separation in both groups, following the challenge tests. Similar to the multivariate case, the number of perturbed bins was less in the high fatigue group, which may support the idea that fatigue hinders the ability to respond to a metabolic challenge, calling for a further comparison between the high and low group post-intervention bin values. For this, the perturbed bins, identified within the high and low fatigue groups when comparing profiles before and after the intervention, were identified, combined and quantified to understand if indicators of the challenge tests could also be associated with the severity of fatigue. Bins with PLS-DA VIP (variable importance in projection) scores exceeding 2, an absolute fold change exceeding 2 and p-values less than or equal to 0.05 (after correcting for multiple testing by controlling the false discovery rate) were selected to be quantified relative to creatinine.

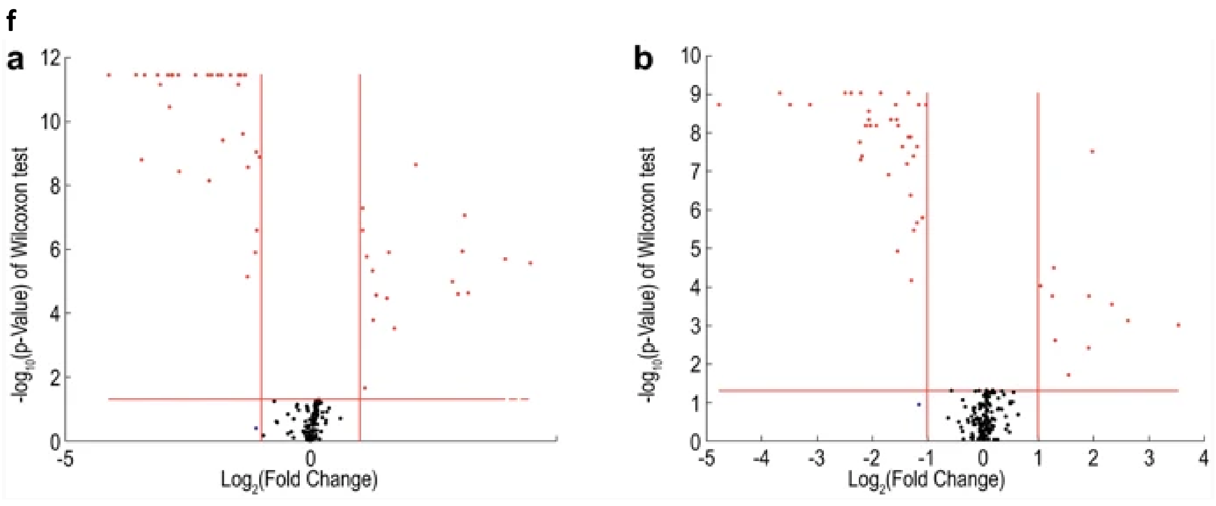

Fig. 3. Statistical assessments of spectral bins indicative of the response following the hepatic challenges. Volcano plot mapped by the scaled fold change and *p*-values for the ^{1}H-NMR bins observed for the low (**a**) and the high (**b**) fatigue groups. Bins with high FC and significant *p*-values among cases are indicated by red dots

Metabolite profiles from a representative cases following the hepatic challenges

Representative ^{1}H-NMR spectra from a single fatigue patient, prior to (black) and following the challenge test (blue), is shown in Fig. 4, scaled according to the creatinine methyl peak at 3.04 ppm. Expanded regions, framed in red in the spectra, are the regions where variables associated with important bins are located. Also, indicated in Fig. 4, are other endogenous urinary metabolites: lactic acid, alanine, citric acid, carnitine, TMAO, and hippuric acid. Regarding the acetaminophen challenge, both the challenging substance (acetaminophen) as well as two conjugation products, acetaminophen-sulphate and acetaminophen-glucuronide, were detected in the urine samples following the challenge. These results firstly indicate that two of the three major phase II conjugation reactions of acetaminophen were functional as detoxification reactions following the challenge: (1) conjugation of glucuronic acid to the hydroxyl group of acetaminophen; (2) sulphation of the phenolic hydroxyl group of acetaminophen. Apparently, the combination of phase I and phase II conjugation between acetaminophen and glutathione, producing *N*-acetyl-p-benzoquinone imine as final product, was not operative in the present experimental challenge. Secondly, no acetyl-salicylic acid was detected in the urine following the known high de-acylation following consumption of this challenge substance. This is related to the coupled de-acylation and glycination of the functional carboxylic group is catalysed by glycine-*N*-acyl transferase (GLYAT). Thus, two metabolites derived from the acetylsalicylic acid challenge could be detected: salicylic acid and salicyluric acid (2-hydroxyhippuric acid).

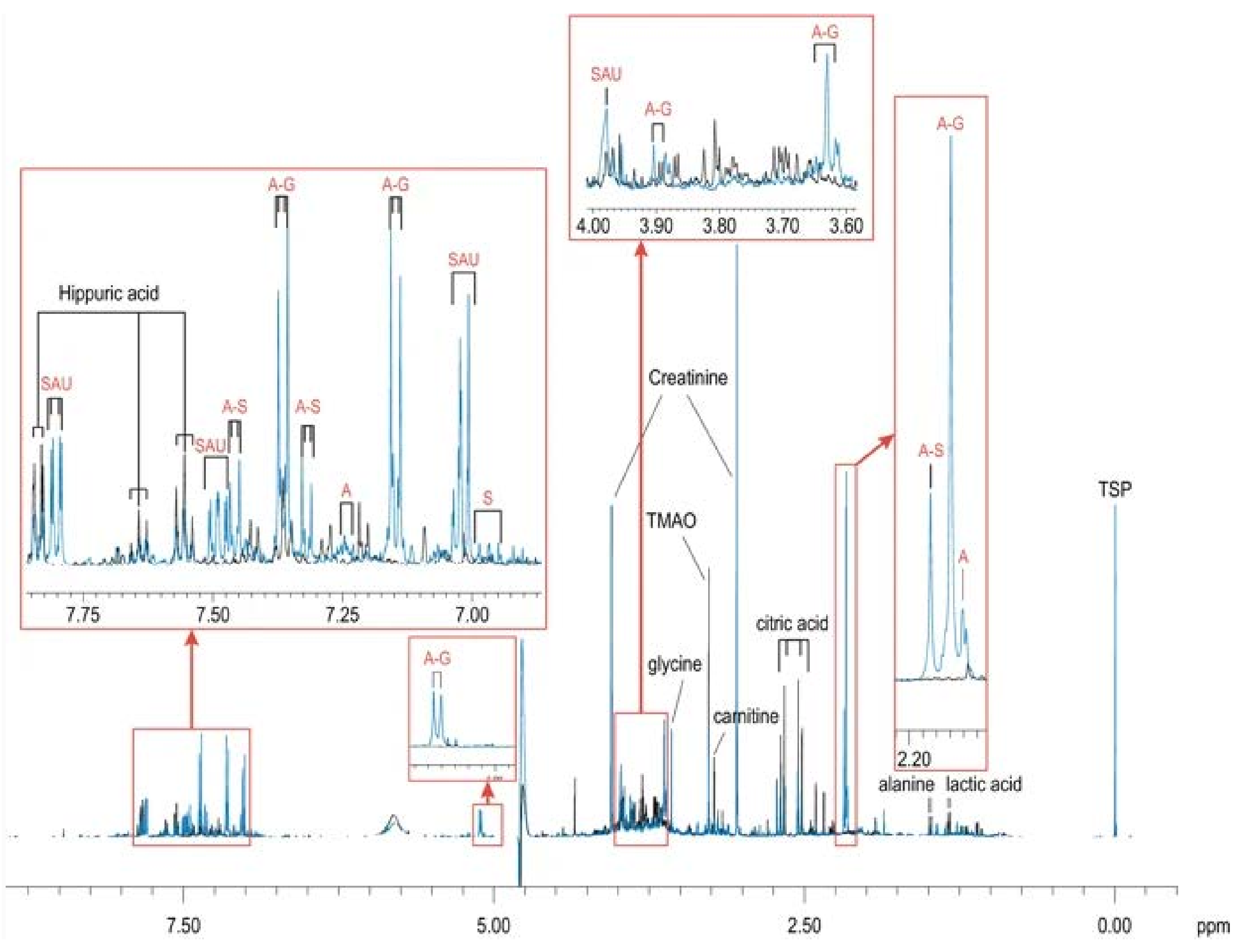

Fig. 4. epresentative ^{1}H-NMR spectra from a single fatigue patient, prior to (black) and following the challenge test (blue), scaled according to the creatinine methyl peak at 3.04 ppm. Expanded regions, framed in red in the spectra, are the bin regions where variables important in projection (VIP) through the supervised PLS-DA are located. VIPs: *A* acetaminophen, *A-G* acetaminophen glucuronide, *A-S* acetaminophen sulfate, *S* salicylic acid, *SAU* salicyluric acid

Quantitative comparison of the two fatigue groups following the challenges

As indicated in Fig. 4, all substances observed following the challenge tests were detected in more than one region in the ^{1}H-NMR spectra. A summary of these areas (ppm values) and detected substances is summarized in Table 1. Apart for these five substances, bins associated with endogenous metabolites, tyrosine, creatine and glucose, were also selected as significantly perturbed during the challenge. We therefore quantified these eight substances for all low and high fatigue cases, followed by statistical analysis of these data.

Table 1 Chemical information and shifts for representative ^{1}H-NMR spectra of CFS patient

Peak assignment	Metabolite name	Chemical structure	HMDB ID	Chemical formula	δ (ppm), multiplicity (J coupling)	Chemical group	Atom (s)
A	Acetaminophen		HMDB0001859	$C_8H_9NO_2$	2.17 s	CH_3	2
					7.25 d (J = ND)	$(CH)_2$	6,10
A-G	Acetaminophen-glucuronide		HMDB0010316	$C_{14}H_{17}NO_8$	2.17 s	CH_3	21
					3.62–3.89 m	$(CH)_4$	2,3,4,5
					5.11 d (J = 7.2 Hz)	CH	6
					7.15 dd (J = 6.8, 9.1 Hz)	$(CH)_2$	16,18
					7.36 dd (J = 6.8, 9.1 Hz)	$(CH)_2$	15,19
A-S	Acetaminophen-sulphate		HMDB0059911	$C_8H_9NO_5S$	2.18 s	CH_3	13
					7.31 dd (J = 6.7, 9.1 Hz)	$(CH)_2$	8,10
					7.45 dd (J = 6.7, 9.1 Hz)	$(CH)_2$	7,11
S	Salicylic acid		HMDB0001895	$C_7H_6O_3$	6.97 m	$(CH)_2$	4,6
SAU	Salicyluric acid		HMDB0000840	$C_9H_9NO_4$	3.96 bs	CH_2	2
					7.01 dd (J = Hz)	$(CH)_2$	10,12
					7.49 m	CH	11
					7.79 dd (J = Hz)	CH	9

Multiplicity: *s* singlet, *d* doublet, *dd* double doublet, *m* multiplet, *bs* broad singlet, *ND* not determined

The outcome of the univariate and multivariate analysis of the potential five diagnostic substances, derived from the challenge tests, are shown in Table 2. Firstly, the PLS-DA for the low and high fatigue groups when comparing substances before and after the challenge tests using the eight substances mentioned above, show complete separation, as expected (Fig. S1). An isolated peak for each identified VIP metabolite was selected, as indicated by the 'reference bin' in Table 2. Each selected peak was integrated and made relative to the creatinine methyl peak at 3.04 ppm, taking into account the number of protons representing each peak. The mean and standard deviations are given in Table 2 as mmol/mol creatinine concentrations.

Table 2 Univariate comparison between high and low fatigue groups

	Metabolite name	Reference bin (max bin VIP)	p-value (effect size)	Low Mean (SD)	High Mean (SD)
Acetaminophen challenge	Acetaminophen	2.15 (7.1)	0.08 (**0.5**)	83.02 (34.4)	100.15 (57.23)
	Acetaminophen glucuronide	2.19 (7.4)	0.08 (**0.66**)	465.18 (163.87)	573.63 (355.05)
	Acetaminophen sulphate	2.17 (7.7)	0.95 ($\leq$0.001)	330.56 (104.19)	330.86 (117.66)
Acetyl-salicylic acid challenge	Salicylic acid	6.95 (4.4)	0.88 (0.12)	49.85 (40.89)	45.10 (25.44)
	Salicyluric acid	7.81 (10.1)	0.06 (0.28)	408.08 (169.07)	455.17 (108.31)
Endogenous	Glucose	5.19 (2.4)	0.06 (0.24)	22.53 (12.15)	25.41 (7.32)
	Creatine	3.03 (2.8)	0.41 (0.22)	208.72 (158.99)	173.42 (142.63)
	Tyrosine	7.23 (3.2)	0.78 (0.07)	83.86 (44.59)	80.69 (39.72)

Bold refers to compounds which were statistically significant at 10% level ($p \leq 0.1$) and of practical relevance (ES$\geq$0.5)

It reports the p-values obtained when comparing high and low fatigue groups across eight compounds of interest, after exposure to the intervention. p-values were obtained using an independent samples t-test (not assuming equal variances) based on transformed data with no correction for multiple testing. Effect sizes were also calculated on untransformed data based on a variation of Cohen's d-value with the low fatigue group as a control. Also reported are means and standard deviations (SD) for each compound for each fatigue group prior to data transformation

The effect of the challenge is thus confirmed in the quantified data. The difference between the mean values of the low and high fatigue groups for these indicators of a challenge were however not statistically significant. Comparing the effect sizes (ES), based on Cohen's d values, indicate to deliver a practical visible effect (ES ≥ 0.5) (Ellis and Steyn 2003) for glucuronide-acetaminophen (ES = 0.7) and acetaminophen (ES = 0.5). Neither of these correlations is of practical relevance ($|r| < 0.5$) and point to no more than a slight but consistent association between variables and their ranks, respectively. No correlation was observed between the biotransformation parameters and the mental fatigue scores. This leads one to conclude that an effect is visible, but greater statistical power is required to detect it.

Discussion

Against the background of the results shown here, what did the metabolomics approach revealed on the applicability of a regimen of challenge tests to objectively assess fatigue?

Firstly, untargeted ^{1}H-NMR metabolomics proved its well-established ability as a global profiling tool to reveal the impact of toxicological interventions (Griffin 2003). However, one of the big hurdles in metabolomics studies is validation, specifically regarding sample size, as recently reviewed by Johnson et al. 2016. Notwithstanding the large cohort of 500+ fatigue patients from which the control (low) and patient (high) fatigue groups were selected, the exclusion criteria resulted in cases for both groups of only 30+ cases. For such a group, the model as shown for example in Fig. 2e, f did not validate well. Within the limitation of validation, here specifically applied on the biotransformation of challenge substances, using data from spectra (Fig. 4) in conjunction with statistical pattern recognition technique (Figs. 2 and 3) did provide a usable tool for profiling of the successive biotransformation metabolites. Biotransformation remains one of the most important defence mechanisms against xenobiotic insult (Xu et al. 2005). Occurring mainly in the liver, it entails initials the absorption of the xenobiotic compound, followed by Phase I functionalization and Phase II conjugation. The products of these reactions are often more hydrophilic and their final excretion from within the cell is facilitated by Phase III (transport mediated) biotransformation.

The present results are the first of its kind on using two xenobiotics in combination in a challenge test on hepatic function in cases of fatigue. The metabolomics information on the acetaminophen challenge clearly supports the existing views on the biotransformation of both xenobiotics. acetaminophen has a half-life of 1.5–3 h after a therapeutic dose, while the major part of the acetaminophen probe is reportedly excreted as the acetaminophen-glucuronide (50–70%) and acetaminophen (25–35%). A small percentage (5–15%) of the administered acetaminophen is converted by a Phase I enzyme (CYP2E1) to NAPQI, a very reactive quinone, which is normally conjugated to glutathione and ultimately excreted as a mercapturate conjugate. This conjugation product was not detected in any of the present cases studied, salicylic acid and salicylic uric acid, derived from each of the acetaminophen and acetyl-salicylic acid challenging substances respectively. The second xenobiotic used, acetylsalicylic acid, is rapidly deacetylated to salicylic acid. The conjugation product of salicylic acid with glycine produces salicyluric acid, the main phase II liver metabolite. The excretion of salicyluric acid normally ranges from 19.8 to 65% of the administered dose. The presence of the both conjugates in the urine of the patients who

were subjected to the challenge tests corresponds with the expected metabolism of the administered of both challenge substances, confirming the value of metabolomics as a tool in the study of a xenobiotic challenges from more than one substance. Furthermore, the observations from the present untargeted metabolomics approach compare well to the complementary technique of targeted metabolomics that was previously described for a biotransformation, using p-aminobenzoic acid as exotoxin (Nortje et al. 2015).

Secondly, the functionality of the biotransformation pathways may be affected by a number of factors: (1) Genetic diversity in the enzymes coding for biotransformation enzymes may affect the ability to up-regulate the production of the enzyme under inductive conditions (van der Sluis et al. 2015; van der Sluis 2018). (2) Conjugation substrates and cofactors are needed for maintaining an effective rate of biotransformation. If depleted, it may negatively impact on biotransformation capacity of both phase I and phase II biotransformation. (3) All of the Phase II reactions are heavily energy dependent and requires ATP for activation of xenobiotica to the corresponding CoA derivative. Although we did not address these factors influencing biotransformation, we observed some qualitative differences between the pre and post-challenge profiles of the low versus high fatigue cases. Both the PCA (Fig. 2a and b) and the PLSDA (Fig. 2c and d) suggest that the low fatigue group responded more to the challenge test than the high fatigue group. Likewise, the volcano plots (Fig. 3) indicated that the number of ^{1}H-NMR-spectral bins with significant fold-changes was less in the high than in the low fatigue group. These observations will be compatible with the view that the high fatigue group suffered from energy depletion due to up-regulated biotransformation related to endogenous or exogenous conditions causative for their symptoms of high fatigue.

Thirdly, although no clear potential biomarker or biosignature for chronic fatigue emerged from this study, the different relative changes in the responses of the individuals comprising the low and high fatigue groups, as well as the quantitative profile of metabolites from the challenges hold distinct promise for follow-up studies. The phase II biotransformation of acetaminophen, as well as the urinary excretion of the parent substance proved to be a practical visible effect considering glucuronide-acetaminophen (ES = 0.7) and acetaminophen (ES = 0.5). Likewise, a somewhat related observation was seen in the excretion of salicyluric acid (VIP = 1.5 p =0.58) in the high fatigue group. This may be further supportive of the suggestion that the high fatigue group react to a higher xenobiotic-load by up-regulating of crucial phase II enzyme systems.

These observations hold promise for further studies aimed at the development of an objective measurement of chronic fatigue. Directives for future research would be (1) the use of a control group without any signs of fatigue (PFS energy and mental ≤ 0.5, with no indications of any mental or clinical conditions), (2) doing a longitudinal analysis of urinary excretion of the diagnostic substances for optimizing an improved endpoint following the challenge and (3) eventually use an alternative experimental design and statistical approaches specifically directed to intervention studies, like ANOVA Simultaneous Component Analysis (or ASCA).

Conclusion

Our view still holds that it seems unlikely that an objective measure, as a verified a single biomarker or biosignature, will be found for fatigue, given the multifactorial nature of these conditions. The renewed emphasis on fatigue conditions may reveal the causes and pathophysiology underlying fatigue. Meanwhile we believe that the outcomes of the present metabolomics study indicates that intervention studies clearly provide for an avenue for an objective way for group differentiation towards a challenge provided for cases of low and high fatigue, which is a corner stone to pave the way towards an objective indicator or marker for fatigue.

Funding

Research funding for the analytical and computational aspects of the project was provided by the Technological Innovation Agency (TIA) of the Department of Science and Technology of South Africa. The views and findings of this investigation are those of the authors and do not reflect or represent any policies of the TIA funder. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Contributions

EE was responsible for the experimental design on fatigue in the cohort of patients, inter-action with the physicians and for the collection of the samples and the biochemical analysis, as well as data collection from the questionnaires. SM did the ^{1}H-NMR analysis and data generation, and provided the quantitative metabolite information on the detoxification challenge in the selected cases. FES did the statistical analysis on the link between the medical and fatigue scales. MvR did all pre-processing of the data and provided the univariate and multivariate analysis of all data. CV provided the input for the clinical aspects of the study. CJR participated in all areas of the research, data analysis and proposed the original outline of the manuscript, to which all authors provided their expert contributions. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

This study complied with all institutional guidelines of the North-West University as stipulated by the South African Guidelines for Good Clinical Practice Ethical Guidelines for Research, as well as the terms of the Declaration of Helsinki of 1975 (as revised in 2013) for investigation of human participants. Ethical approval was obtained from the Health Research Ethics Committee (HREC) of the North-West University (NWU-00102-12-A1). All written consent was obtained based upon informed decision from all participants. An example of the informed consent form can be found in SI.

References

Clayton, E. W. (2015). Beyond myalgic encephalomyelitis/chronic fatigue syndrome: An IOM report on redefining an illness. *JAMA, 313*(11), 1101–1102.

Cordts, P. R., Kaminski, M. V., Raju, S., Clark, M. R., & Woo, K. M. (2001). Could gut-liver function derangements cause chronic venous insufficiency? *Vascular Surgery, 35*(2), 107–113.

Ellis, S. M., & Steyn, H. S. (2003). Practical significance (effect sizes) versus or in combination with statistical significance (p-values). *Management Dynamics, 12*(4), 51–53.

Erasmus, E., Steffens, F. E., Van Reenen, M., Vorster, B. C., & Reinecke, C. J. (2019). Biotransformation profiles from a cohort of chronic fatigue women in response to a hepatic detoxification challenge. *PLoS ONE, 14,* 1–22.

Fourie, N. H., Wang, D., Abey, S. K., Sherwin, L. B., Joseph, P. V., Rahim-Williams, B., et al. (2016). The microbiome of the oral mucosa in irritable bowel syndrome. *Gut Microbes, 7*(4), 286–301.

Freidin, M. B., Wells, H. R., Potter, T., Livshits, G., Menni, C., & Williams, F. M. (2018). Metabolomic markers of fatigue: Association between circulating metabolome and fatigue in women with chronic widespread pain. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1864*(2), 601–606.

Griffin, J. L. (2003). Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Current Opinion in Chemical Biology, 7*(5), 648–654.

Hackshaw, K. V., Aykas, D. P., Sigurdson, G. T., Pujolras, M. P., Madiai, F., Yu, L., et al. (2018). Metabolic fingerprinting for diagnosis of fibromyalgia and other rheumatologic disorders. *Journal of Biological Chemistry, 294*(7), 2555–2568.

Hadrévi, J., Björklund, M., Kosek, E., Hällgren, S., Antti, H., Fahlström, M., et al. (2015). Systemic differences in serum metabolome: A cross sectional comparison of women with localised and widespread pain and controls. *Scientific Reports, 5,* 15925.

Irwin, C., van Reenen, M., Mason, S., Mienie, L. J., Westerhuis, J. A., & Reinecke, C. J. (2016). Contribution towards a metabolite profile of the detoxification of benzoic acid through glycine conjugation: An intervention study. *PLoS ONE, 11*(12), e0167309.

Irwin, C., van Reenen, M., Mason, S., Wevers, R. A., Westerhuis, J. A., & Reinecke, C. J. (2018). The ^{1}H-NMR-based metabolite profile of acute alcohol consumption: A metabolomics intervention study. *PLoS ONE, 13*(5), e0196850.

Johnson, C. H., Ivanisevic, J., & Siuzdak, G. (2016). Metabolomics: Beyond biomarkers and towards mechanisms. *Nature Reviews Molecular Cell Biology, 17*(7), 451–459.

Malatji, B. G., Meyer, H., Mason, S., Engelke, U. F., Wevers, R. A., Reenen, M., et al. (2017). A diagnostic biomarker profile for fibromyalgia syndrome based on an NMR metabolomics study of selected patients and controls. *BMC Neurology, 17*(1), 88.

Nortje, C., Jansen van Rensburg, P., Cooke, C., & Erasmus, E. (2015). The simultaneous detection and quantification of p-aminobenzoic acid and its phase 2 biotransformation metabolites in human urine using LC–MS/MS. *Bioanalysis, 7*(10), 1211–1224.

Piper, B. F., Dibble, S. L., Dodd, M. J., Weiss, M. C., Slaughter, R. E., & Paul, S. M. (1998). The revised Piper Fatigue Scale: Psychometric evaluation in women with breast cancer. *Oncology Nursing Forum, 25,* 677–684.

Ponnusamy, K., Choi, J. N., Kim, J., Lee, S. Y., & Lee, C. H. (2011). Microbial community and metabolomic comparison of irritable bowel syndrome faeces. *Journal of Medical Microbiology, 60*(6), 817–827.

van der Sluis, R. (2018). Analyses of the genetic diversity and protein expression variation of the acyl: CoA medium-chain ligases, ACSM2A and ACSM2B. *Molecular Genetics and Genomics, 293*(5), 1279–1292.

van der Sluis, R., Badenhorst, C. P., Erasmus, E., Van Dyk, E., Van der Westhuizen, F. H., & Van Dijk, A. A. (2015). Conservation of the coding regions of the glycine N-acyltransferase gene further suggests that glycine conjugation is an essential detoxification pathway. *Gene, 571*(1), 126–134.

Westerhuis, J. A., Hoefsloot, H. C., Smit, S., Vis, D. J., Smilde, A. K., van Velzen, E. J., et al. (2008). Assessment of PLSDA cross validation. *Metabolomics, 4*(1), 81–89.

Wittwer, J., Rubio-Aliaga, I., Hoeft, B., Bendik, I., Weber, P., & Daniel, H. (2011). Nutrigenomics in human intervention studies: Current status, lessons learned and future perspectives. *Molecular Nutrition & Food Research, 55*(3), 341–358.

Xu, C., Li, C. Y. T., & Kong, A. N. T. (2005). Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Archives of Pharmacal Research, 28*(3), 249–268.