

A metabolomics investigation on experimental interventions of acute alcohol consumption

C Irwin

 **orcid.org 0000-0003-4630-8468**

Thesis submitted in fulfilment of the requirements for the degree
Doctor of Philosophy in Biochemistry at the North-West
University

Promoter:	Prof CJ Reinecke
Co-promoter:	Dr SW Mason
Assistant Promoter:	Prof JA Westerhuis

ACKNOWLEDGEMENTS

Appreciation, thanks and acknowledgement goes out to everyone who participated in this project and supported me on the long (and sometimes very difficult) road of my Ph.D. study:

Prof. Carools Reinecke, for your help with the conceptualization of the study and the development of the methodological design, as well as for obtaining research funds and ethical approval for the project, and for supervising the planning and execution of the investigation. But more important than any of that, thank you for your constant support, guidance and motivation throughout the many years I have known you, studied under you and learnt from you. You are a true academic inspiration!

Prof. Johan Westerhuis, for your help with conceptualizing the study and developing the methodological design, as well as for your help with conceptualizing the complementary techniques required for the statistical analysis of the data.

Dr. Shayne Mason, for your help with the generation of the ^1H -NMR data and assessment of the spectral analyses, as well as for your guidance throughout my Ph.D. journey.

Mari van Reenen, for providing me with your bioinformatics expertise, for your help with conceptualizing the complementary techniques required for the statistical analysis of the data, as well as for performing all the statistical and computational analyses of the original and curated data.

Prof. Japie Mienie, for allowing me to perform my GC–MS analyses in your lab, for your help with the GC–MS data generation, data curation and analysis, as well as for your participation throughout the investigation as an expert on metabolism, and in the modelling of the proposed metabolic outcomes of the study. Without your valuable insights into metabolic processes the biological interpretation of the results would not have been possible.

Prof. Ron Wevers, for your participation as an expert on metabolism, and in the modelling of the proposed metabolic outcomes of the study.

Acknowledgements also go out to the service providers, academic institutions, staff and individuals affiliated with this study and/or thesis:

- The Centre for Human Metabolomics, Faculty of Natural and Agricultural Sciences, North-West University (Potchefstroom Campus), South Africa.
- Everyone at Potchefstroom Laboratory for Inborn Errors of Metabolism (PLIEM), a division of Biochemistry, North-West University (Potchefstroom Campus), South Africa, who contributed in some way.

- The Technology and Innovation Agency (TIA) of South Africa for funding the project.
- All the co-authors who contributed towards the three publications for your expert input in the development and writing of the manuscripts.
- Dr. Graham Baker and Dr. Elisabeth Lickindorf at Kerlick Editorial and Research Solutions for your superb editing of the three publications, as well as this thesis.
- Prof. Casper Lessing for expertly editing my large number of references.
- The 24 individuals who volunteered to take part in this study, as well as the physician, Dr. Chris Vorster, who supervised the experimental proceedings and sample collection.

My parents, family and friends, thank you for your continued support, words of encouragement and understanding during this journey, even though I was sometimes difficult to cope with.

Finally, my incredibly amazing fiancé — thank you, thank you, thank you, a million times over! Words simply cannot describe what your never-ending support means to me. Your love, compassion and understanding (even when I am at my worst) continues to amaze me. Without you I would not have had the strength and will to carry on and complete this challenging venture. I love you with all my heart!

ABSTRACT

This thesis, titled: “*A metabolomics investigation on experimental interventions of acute alcohol consumption*”, deals with a current topic of global interest, namely, alcohol use and abuse. Alcohol abuse is associated with many serious, and even detrimental, health, social and economic consequences, and is one of the world’s leading risk factors for disability, morbidity and mortality. For these reasons it is a topic of growing concern in developing, as well as developed, countries.

Alcohol is metabolized mainly in the liver by two nicotinamide adenine dinucleotide (NAD)-dependent enzymes — alcohol dehydrogenase and, subsequently, aldehyde dehydrogenase. In both of these reactions oxidized NAD (NAD^+) is reduced to NADH, which increases the $\text{NADH}:\text{NAD}^+$ ratio in hepatocytes. This ratio controls the activity of several key metabolic enzymes and the direction of many reversible metabolic reactions, and its disruption is known to result in perturbations of various metabolic pathways.

Various studies examining the effects of, and diseases related to, chronic alcohol abuse have been performed in the last few decades. However, to date, no comprehensive metabolomics study on the effects of acute alcohol consumption has been done. Thus, with the guidance of experts in the fields of metabolism, metabolomics and biostatistics, the first extensive, multidisciplinary metabolomics cross-over intervention study into the effects of acute alcohol consumption on the urinary metabolite profiles of healthy, young males was designed, and is presented in this thesis.

The study consisted of analysing urine samples, collected from experimental participants over a defined period of time following four interventions, on two different analytical platforms — proton nuclear magnetic resonance (^1H -NMR) spectroscopy as an untargeted approach, and gas chromatography–mass spectrometry (GC–MS) as a semi-targeted approach. The results from these investigations demonstrated the power of applying metabolomics to this area of research and provided the opportunity to obtain a holistic view of the urinary metabolic profile resulting from acute alcohol consumption.

From both of these approaches a list of metabolites perturbed by acute alcohol consumption could be compiled with the use of statistical analyses. Various metabolic pathways were seen to be disrupted, most of them due to the known alcohol-induced increased $\text{NADH}:\text{NAD}^+$ ratio. Additionally, two urinary metabolites — sorbitol, from the ^1H -NMR analysis, and 2-hydroxyisobutyric acid, from the GC–MS investigation — not previously known to be associated with the consequences of acute alcohol consumption were identified in the metabolic profiles of the experimental participants following acute alcohol consumption. These novel findings could

possibly be used as a basis for determining biomarkers of acute alcohol consumption, which could have various health, economic and legal benefits.

This thesis, the eventual product of a skilfully designed and diligently carried out scientific study, is compiled and presented in article format as per the requirements of North-West University. The scientific contributions made during this study to the existing alcohol-related scientific knowledge resulted in three publications. Two (1 and 3) have already been published, and one (2) has been accepted for publication.

1. 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. doi:10.1371/journal.pone.0167309.
2. Irwin, C., Van Reenen, M., Mason, S., Mienie, L.J., Wevers, R.A., Westerhuis, J.A. & Reinecke, C.J. The ^1H -NMR-based metabolite profile of acute alcohol consumption: a metabolomics intervention study.
3. Irwin, C., Mienie, L.J., Wevers, R.A., Mason, S., Westerhuis, J.A., Van Reenen, M. & Reinecke, C.J. 2018. GC–MS-based urinary organic acid profiling reveals multiple dysregulated metabolic pathways following experimental acute alcohol consumption. *Scientific Reports*, 8:5775. doi:10.1038/s41598-018-24128-1.

Keywords: acute alcohol (ethanol); cross-over intervention; nicotinamide adenine dinucleotide (NAD); metabolomics; urine; proton nuclear magnetic resonance (^1H -NMR) spectroscopy; gas chromatography–mass spectrometry (GC–MS); statistical analyses.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	I
ABSTRACT	III
PART I AN OVERVIEW OF ETHANOL METABOLISM	1
CHAPTER 1 INTRODUCTION	1
1.1 Background	1
1.2 Structure of this thesis.....	5
REFERENCES.....	8
CHAPTER 2 ETHANOL METABOLISM	10
2.1 The metabolism of ethanol	10
2.2 Nicotinamide adenine dinucleotide (NAD)	12
2.3 The effect of acute alcohol consumption on metabolism.....	14
2.4 The effect of chronic alcohol consumption on metabolism	19
REFERENCES.....	24
PART II NMR METABOLOMICS OF ACUTE ALCOHOL CONSUMPTION	29
CHAPTER 3 METABOLOMICS AND ALCOHOLISM.....	29
3.1 What is metabolomics?.....	29
3.2 Metabolomics and chronic alcohol consumption	36
3.2.1 Hyphenated MS studies.....	36
3.2.2 NMR studies	39
3.2.2.1 Experimental animal studies	39

3.2.2.2	Global metabolite profile of chronic alcoholism	40
3.2.2.3	New insights from chronic alcoholism studies	43
3.3	Metabolomics and acute alcohol consumption.....	44
REFERENCES.....		48
CHAPTER 4 NMR METABOLOMICS OF ACUTE ALCOHOL CONSUMPTION		53
4.1	Background on intervention studies.....	53
4.2	Experimental design for an intervention study	54
4.3	Perspectives on NMR technology as a mode for the generation of metabolomics data	59
4.3.1	Sample preparation	59
4.3.2	¹ H-NMR analysis	60
4.3.3	Quality assurance.....	60
4.4	An intervention study on the detoxification of benzoic acid	62
4.4.1	Background	62
4.4.2	Contribution towards a metabolite profile of the detoxification of benzoic acid through glycine conjugation: an intervention study (Irwin <i>et al.</i> , 2016)	62
4.5	An intervention study on acute alcohol consumption.....	99
4.5.1	Background	99
4.5.2	The ¹ H-NMR-based metabolite profile of acute alcohol consumption: a metabolomics intervention study (Irwin <i>et al.</i> , 2018 – Accepted).....	100
REFERENCES.....		137
PART III	GC-MS METABOLOMICS OF ACUTE ALCOHOL CONSUMPTION.....	141
CHAPTER 5 GC-MS METABOLOMICS OF ACUTE ALCOHOL CONSUMPTION.....		141

5.1	Metabolism: The main root of GC-based metabolomics.....	141
5.2	Sources of variation and quality control in GC-metabolomics	146
5.3	From GC–MS data to biology.....	156
5.4	Experimental and measurement design for an interventional GC–MS study.....	160
5.4.1	Background	160
5.4.2	GC–MS measurement design.....	161
5.4.3	Extraction and derivatization of organic acids	164
5.5	Metabolite identification and classification	165
5.6	Additional experiences with sources of variation	167
5.6.1	Artifact formation	168
5.6.2	Repeatability of QC sample analyses	170
5.6.3	Laboratory logistics.....	171
5.6.4	Problem solving	172
5.7	Outcomes of a GC–MS metabolomics study on acute alcohol consumption (Irwin <i>et al.</i>, 2018).....	173
REFERENCES.....		206
PART IV	NEW INSIGHTS ON ACUTE ALCOHOL CONSUMPTION	216
CHAPTER 6 DISCUSSION.....		216
6.1	Reflecting on the aims of my study.....	216
6.1.1	The first aim of the study: Provide a holistic view of the comprehensive and complex metabolic consequences that occur due to the consumption of an acute dose of alcohol.....	216

6.1.2	The second aim of the study: Obtain new insights into the biochemical processes and pathways affected by acute alcohol consumption, not observed through the traditional biochemical investigations.....	218
6.2	Lessons learned through this study	221
6.3	Future prospects	225
6.3.1	2-Hydroxyisobutyric acid.....	225
6.3.2	Biomarker identification	225
6.3.3	Genetic polymorphisms	227
6.3.3.1	What is a polymorphism?	227
6.3.3.2	The effect of polymorphisms on alcohol consumption.....	227
6.3.3.3	Future studies on acute alcohol consumption	231
	REFERENCES.....	234
	ANNEXURE	238
1	Ethical approval certificate	238
2	Informed consent form.....	239
3	Copyright licences of journals.....	242
4	Instructions to authors from journals	244
5	Permission and contributions of co-authors.....	277
6	Copyright licences for figures	280

LIST OF TABLES

Table 3-1:	Examples of recent metabolomics studies (performed using various LC–MS techniques) on the effect of chronic alcohol consumption on metabolism, as well as of chronic disease states related to chronic alcohol consumption.	38
Table 5-1:	Sampling procedures	149
Table 5-2:	Preparation of samples for GC–MS analysis	150
Table 5-3:	GC separation of sample compounds	151
Table 5-4:	MS analysis of separated compounds.....	153
Table 5-5:	Analysis of GC–MS spectra	154
Table 5-6:	Analysing samples in multiple batches — batch effect	155
Table 5-7:	Data analysis and processing	155
Table 6-1:	Details of the various ADH and ALDH genes related to alcohol metabolism	230

LIST OF FIGURES

Figure 1-1:	Original notes on the experimental design for a metabolomics alcohol intervention study.	4
Figure 2-1:	Schematic representation of the conversion between NAD^+ and NADH during redox reactions [adapted from Rye <i>et al.</i> , 2016].	13
Figure 2-2:	Summarized model of the global perturbations induced by excessive NADH (increased $\text{NADH}:\text{NAD}^+$ ratio) as a result of ethanol oxidation.....	15
Figure 3-1:	Metabolomics workflow [adapted with permission from Hendriks <i>et al.</i> , 2011].....	30
Figure 3-2:	OPLS score plot and spectra of chronic liver failure (CLF) and acute-on-chronic liver failure (ACLF) patients [reproduced from Amathieu <i>et al.</i> , 2014].....	41
Figure 4-1:	Structure of the basic experimental design for all experimental subjects. <i>K</i> represents the interventions in the study.....	57
Figure 4-2:	Bar graph illustrating the CV% values of the QC samples analysed using NMR spectroscopy.....	61
Figure 5-1:	Number of NMR metabolomics and MS metabolomics publications [reproduced with permission from Emwas, 2015].	142
Figure 5-2:	ERNDIM certificate of participation and performance of the PLIEM laboratory of North-West University.....	159
Figure 5-3:	Representation of all elements of the experimental design.....	160
Figure 5-4:	Measurement design for batches..	162
Figure 5-5:	PCA scores plot of the QC samples (blue) and experimental samples (pink) from the 12 batches selected to be quantified..	173
Figure 6-1:	Revised and expanded version of the model of the global metabolic perturbations induced by the alcohol-induced increased $\text{NADH}:\text{NAD}^+$ ratio.....	220

Figure 6-2:	Illustration of two main methods of undertaking a study such as this one with regard to GC–MS sample preparation and analysis, and interpretation of the results.....	221
-------------	---	-----

PART I AN OVERVIEW OF ETHANOL METABOLISM

CHAPTER 1 INTRODUCTION

“Every new beginning comes from some other beginning’s end.”

These words from the Roman philosopher Seneca the Younger are no more applicable to any other field than they are to science. Indeed, they effectively encapsulate the essence of science. New ideas and research endeavours often arise from previous studies’ findings, results and hypotheses, and this is how science and knowledge grow. This was essentially also the case with my study, since the inspiration behind it was gained from a pilot study on acute alcohol consumption, conducted in our laboratory (Mason, 2010).

1.1 Background

Alcohol is a widely available and used psychoactive drug, and the excessive consumption thereof is a current, relevant and exceedingly important research topic. Alcohol use and abuse is becoming increasingly problematic in developing as well as developed countries (Manna *et al.*, 2011). In fact, the 2014 status report by the World Health Organization (WHO) stated that alcohol remains one of the world’s leading risk factors for disability, morbidity and mortality — 5.9% of all deaths worldwide are attributable to alcohol consumption, exceeding the deaths from HIV/AIDS (2.8%), violence (0.9%) or tuberculosis (1.7%) (WHO, 2014). Acute and chronic alcohol abuse is associated with an extensive variety of medical, psychological, social, personal, legal, and economic consequences (Gunzerath *et al.*, 2011; Wurst *et al.*, 2005), including traffic accidents and fatalities, physical injuries, depression, suicide, anxiety, fights, alcohol poisoning, various criminal offences, and the development of many acute and chronic diseases (O’Hare, 1997). In fact, the tenth edition of the *International Classification of Disease* lists at least 25 chronic conditions that are entirely attributable to alcohol (Shield *et al.*, 2013). Alcohol consumption is also a risk factor in certain cancers, some tumours, numerous cardiovascular and digestive diseases, and many neuropsychiatric conditions.

There exists tremendous variation in drinking patterns among alcohol consumers, leading to the definition of distinct modes of drinking based on the quantity, frequency, and duration of alcohol consumption:

- Males who consume up to 2 drinks per day are defined as moderate (or low-level) drinkers.

- Acute alcohol consumption (the rapid ingestion of alcoholic beverages) and binge or “risky” drinking (drinking too much too fast) is defined by the US National Institute of Alcohol Abuse and Alcoholism (NIAAA) as the consumption of 5 or more drinks by males within 2 hours, resulting in a blood alcohol concentration of 0.08 g% or above (NIAAA; Roache *et al.*, 2015).
- Chronic or excessive alcohol consumption in males is defined as the consumption of more than 5 drinks per day over a long period of time (Zakhari & Li, 2007).

In each of these cases a standard drink contains approximately 12.5 g of ethanol, and is defined by the US Department of Agriculture Dietary Guidelines as 360 mL of beer or wine cooler, 150 mL of wine, or 45 mL of 80-proof distilled spirits (Zakhari & Li, 2007). Moderate drinking does not appear to be associated with increased health risks when compared to abstaining from alcohol consumption. In fact, studies have shown that moderate drinking may be associated with various health benefits (Hines & Rimm, 2001). The same is, however, not true for acute and chronic alcohol consumption, both of which are associated with serious negative health consequences.

The existing research and encyclopaedic information available on the clinical consequences and metabolic perturbations associated with acute and chronic alcohol consumption is, however, predominantly based on traditional methodologies, the focus being on particular/selected perturbations or metabolic pathways, depending on the research question under investigation (Chapter 2). Alcohol consumption, however, has a disruptive effect on interrelated metabolic pathways within the human body, which results in changes in the concentrations of various metabolites associated with these pathways. The concentrations of these metabolites are often only very slightly different from their normal physiological values. Also, these metabolites originate from various widespread, often minor, metabolic pathways. These two factors complicate the isolation and identification of these metabolites when using traditional methods.

Metabolomics approaches have been used to examine various physiological processes, and have been applied to disease diagnosis, drug discovery, nutrition studies, and toxicological investigations (Li *et al.*, 2011). More recently, several studies have used a metabolomics approach to investigate the effect of chronic alcohol consumption on metabolism (Chapter 3). These studies show significant promise for disease profiling and biomarker identification in conditions associated with chronic alcohol consumption, and illustrate the value of a metabolomics approach to this area of research. Specifically, the examples of studies using nuclear magnetic resonance (NMR) spectroscopic-based metabolite profiling clearly support our view that an NMR-based approach would be a good first analytical method to use in my study on acute alcohol consumption.

A gas chromatographic–mass spectrometric (GC–MS)-based metabolomics approach would be equally applicable and valuable to studies on the effect of acute alcohol consumption on metabolism, since, to date, only one pioneering exploratory GC–MS study on acute alcohol consumption has been reported in the field of metabolomics (Mason, 2010). The study by Mason, however, presented a number of important observations, including the identification of changes in the concentrations of several metabolites as a consequence of acute alcohol consumption. The much higher sensitivity for metabolites present in low concentrations in biological samples makes GC–MS an essential complement to the NMR approach, which, on the other hand, offers a higher specificity over a broad spectrum of metabolites.

The observations and results from the study by Mason (2010) sparked my interest in the metabolic effect of acute alcohol consumption, and it quickly became evident that there are still many gaps in our knowledge of this topic. The application of metabolomics to intervention studies greatly enhances the holistic understanding of the effects of consumed substances on metabolic pathways (Gibney *et al.*, 2005; Wishart, 2008). Therefore, since it seemed promising that metabolomics could be used to fill some of these knowledge gaps, I envisaged using a metabolomics approach to further elucidate the effect of an acute dose of alcohol on the metabolism and urinary metabolic profiles of healthy, moderate-drinking young men. Financial support for a basic study on the metabolomics of alcohol consumption (including the effect of nicotinamide adenine dinucleotide (NAD), a commercial product used to alleviate the effect of alcoholism) was obtained from the Technological Innovation Agency (TIA) in 2012, from where the development and planning of my M.Sc. study began.

In order to determine as reliably, and accurately, as possible the metabolic effect of an intervention, and to minimize as far as possible experimental sources of variation, metabolomics studies require the designing and execution of a robust and carefully controlled standardized analytical and experimental protocol. In order to address the complex issue of acute alcohol consumption most effectively, a novel intervention experimental study design was developed between my promoter and co-promoter in consultation with Professor. Age Smilde of the University of Amsterdam, who is renowned for his views on intervention studies (Hendriks *et al.*, 2011; Jansen *et al.*, 2012; Van Velzen *et al.*, 2008). The conceptual design that emanated from a discussion on 19 June 2012 in Amsterdam (Figure 1–1) formed the basis for the experimental and measurement designs used in my investigation, which were modified for the NMR and GC–MS approaches, as will be shown in the respective sections that follow (Chapters 4 and 5). The application of novel statistical methods of analysis to my study illustrates the multidisciplinary approach and close collaboration that are required between biochemists and statisticians in order to interpret experimental data most accurately and reliably. Thus, the conceptual design for the intervention study also opened the opportunity for a Ph.D. study in Statistics, parallel to

broadly formulated, this study was novel both in its experimental design as well as it being the first metabolomics study on acute alcohol consumption.

1.2 Structure of this thesis

This Ph.D. project, like many other advanced student investigations, was marked with several setbacks. However, the realisation that negative results are not worthless, nor a waste of time, but rather opportunities to learn, ask questions, and eventually enhance one's understanding of the problem at hand, kept me motivated. Fortunately, there were also many triumphs, interesting breakthroughs and promising results, all of which eventually culminated in the writing of this thesis, which I chose to present in four complementary parts:

Part I provides the background for the study as presented above (Chapter 1), as well as a literature review in which the pathways of alcohol metabolism are discussed, and the currently known metabolic effects of acute and chronic alcohol consumption are presented (Chapter 2). The reason for, and the aims of, my study are also further motivated here.

In part II, a brief overview of metabolomics and its use in alcohol-related studies is given, with the focus placed on NMR studies (Chapter 3). Chapter 4 describes the full experimental design for the NMR and GC–MS aspects of the investigation, followed by a brief outline on the basics of NMR spectroscopy required for my study. Chapter 4 also includes two publications in which the results obtained from applying untargeted NMR metabolomics to the collected experimental samples are presented. The first of these publications reveals how the experimental and measurement design was modified for the NMR analysis, and describes the experimental procedures followed for this part of the study.

Part III (Chapter 5) discusses the development and application of GC–MS technology, with metabolism and metabolic defects described as a main root of GC-based metabolomics. Then, the basic analytical procedure, pitfalls and potential sources of variation associated with GC–MS metabolomics, as well as the quality control procedures necessary to account for these problems, are presented. This is followed by a section describing the value of proper biological interpretation of data, a comprehensive description of the measurement design as it was modified for the GC–MS part of the study, a description of how the identified metabolites were classified, and a discussion of some relevant complications experienced with the use of GC–MS. This chapter ends with a publication in which the results obtained from applying semi-targeted GC–MS metabolomics to the collected experimental samples are presented. The procedures for organic acid extraction and derivatization, GC–MS analysis, and variable identification are also given as part of this publication.

Part IV (Chapter 6) is a general discussion on the overall findings, results and insights gained from this investigation. This chapter ends this thesis with future prospects and research ideas on the topic of acute alcohol consumption.

The results and knowledge gained from this investigation were presented as peer-reviewed contributions in the form of:

- Two presentations at international conferences:
 1. A poster, titled “*Biotransformation as a determinant context to explore the exposome*”, presented by myself at the 11th International Conference of the Metabolomics Society in San Francisco, California, USA (29 June–2 July 2015).
 2. An oral presentation titled “*Searching for metabolic changes in human nutritional intervention studies*” partly based on my results and presented by Prof. Johan Westerhuis at the Advances in NMR and MS based metabolomics congress in Padova, Italy (14–16 November 2017).
- Three full research papers:
 1. 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. doi:10.1371/journal.pone.0167309.
 2. Irwin, C., Van Reenen, M., Mason, S., Mienie, L.J., Wevers, R.A., Westerhuis, J.A. & Reinecke, C.J. 2018. The ¹H-NMR-based metabolite profile of acute alcohol consumption: a metabolomics intervention study. *PLOS ONE* [Accepted].
 3. Irwin, C., Mienie, L.J., Wevers, R.A., Mason, S., Westerhuis, J.A., Van Reenen, M. & Reinecke, C.J. 2018. GC–MS-based urinary organic acid profiling reveals multiple dysregulated metabolic pathways following experimental acute alcohol consumption. *Scientific Reports*, 8:5775. doi:10.1038/s41598-018-24128-1.

This thesis is compiled and presented in article format as per the requirements of NWU. The references for the articles are listed at the end of each article in the required format of the specific journals. The references occurring in the non-article sections of chapters are listed at the end of each chapter. At the end of this thesis is the annexure, which includes copies of the ethical approval, the informed consent form for the intervention study, copyright licences of the different journals, the instructions to authors from the different journals, and permission and specific contributions of the various co-authors. These addenda are as prescribed by NWU for a thesis including scientific papers.

Upon reflecting on the outline and outcomes of my study, I consider science to be a never-ending journey towards greater understanding. I think the physicist Lawrence Krauss summed it up perfectly by saying:

“The beauty of science is that it does not claim to know the answers before it asks the questions. There is nothing wrong with not knowing. It means there is more to learn, and as I have said before, ignorance bothers me far less than the illusion of knowledge.”

REFERENCES

- Gibney, M.J., Walsh, M., Brennan, L., Roche, H.M., German, B. & Van Ommen, B. 2005. Metabolomics in human nutrition: opportunities and challenges. *American journal of clinical nutrition*, 82(3):497–503.
- Gunzerath, L., Hewitt, B.G., Li, T-K. & Warren, K.R. 2011. Alcohol research: past, present, and future. *Annals of the New York Academy of Sciences*, 1216(1):1–23.
- Hendriks, M.M.W.B., Van Eeuwijk, F.A., Jellema, R.H., Westerhuis, J.A., Reijmers, T.H., Hoefsloot, H.C.J. & Smilde, A.K. 2011. Data-processing strategies for metabolomics studies. *Trends in analytical chemistry*, 30(10):1685–1698.
- Hines, L.M. & Rimm, E.B. 2001. Moderate alcohol consumption and coronary heart disease: a review. *Postgraduate medical journal*, 77:747–752.
- Jansen, J.J., Szymańska, E., Hoefsloot, H.C., Jacobs, D.M., Strassburg, K. & Smilde, A.K. 2012. Between metabolite relationships: an essential aspect of metabolic change. *Metabolomics*, 8(3):422–432.
- Li, S., Liu, H., Jin, Y., Lind, S., Caic, Z. & Jiang, Y. 2011. Metabolomics study of alcohol-induced liver injury and hepatocellular carcinoma xenografts in mice. *Journal of chromatography B*, 879(24):2369–2375.
- Manna, S.K., Patterson, A.D., Yang, Q., Krausz, K.W., Ilde, J.R., Fornace Jr., A.J. & Gonzalez, F.J. 2011. UPLC–MS-based urine metabolomics reveals indole-3-lactic acid and phenyllactic acid as conserved biomarkers for alcohol-induced liver disease in the *Ppara*-null mouse model. *Journal of proteome research*, 10(9):4120–4133.
- Mason, S.W. 2010. The metabolomics of acute alcohol abuse. Potchefstroom: NWU. (Dissertation – M.Sc.).
- NIAAA (National Institute of Alcohol Abuse and Alcoholism). s.a. Drinking levels defined. <https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking> Date of access: 28 Aug. 2017.
- O'Hare, T. 1997. Measuring excessive alcohol use in college drinking contexts: the drinking context scale. *Addictive behaviors*, 22(4):469–477.

Roache, J.D., Karns, T.E., Hill-Kapturczak, N., Mullen, J., Liang, Y., Lamb, R.J. & Dougherty, D.M. 2015. Using transdermal alcohol monitoring to detect low-level drinking. *Alcoholism: clinical and experimental research*, 39(7):1120–1127.

Shield, K.D., Parry, C. & Rehm, J. 2013. Chronic diseases and conditions related to alcohol use. *Alcohol research: current reviews*, 35(2):155–173.

Van Velzen, E.J., Westerhuis, J.A., Van Duynhoven, J.P., Van Dorsten, F.A., Hoefsloot, H.C., Jacobs, D.M., Smit, S., Draijer, R., Kroner, C.I. & Smilde, A.K. 2008. Multilevel data analysis of a crossover designed human nutritional intervention study. *Journal of proteome research*, 7(10):4483–4491.

WHO (World Health Organization). 2014. Global status report on alcohol and health 2014. http://www.who.int/substance_abuse/publications/global_alcohol_report/en/ Date of access: 24 Aug. 2017.

Wishart, D.S. 2008. Metabolomics: applications to food science and nutrition research. *Trends in food science & technology*, 19(9):482–493.

Wurst, F.M., Alling, C., Aradottir, S., Pragst, F., Allen, J.P., Weinmann, W., Marmillot, P., Ghosh, P., Lakshman, R., Skipper, G.E., Neumann, T., Spies, C., Javors, M., Johnson, B.A., Ait-Daoud, N., Akhtar, F., Roache, J.D. & Litten, R. 2005. Emerging biomarkers: new directions and clinical applications. *Alcoholism: clinical and experimental research*, 29(3):465–473.

Zakhari, S. & Li, T-K. 2007. Determinants of alcohol use and abuse: impact of quantity and frequency patterns on liver disease. *Hepatology*, 46(6):2032–2039.

CHAPTER 2 ETHANOL METABOLISM

Alcohol consumption is a common and legal practice in most countries around the world, and on the rise in developing, as well as in developed, countries (Manna *et al.*, 2011). It is one of the most frequently consumed beverages at social gatherings, is used excessively by many individuals and has become the most frequently abused drug throughout the world (Lieber, 1995). Dr Charles Lieber was a pioneer of modern research on alcohol and alcohol-induced liver damage. He made many significant discoveries and contributions to this field, including recognizing that ethanol itself is a hepatotoxin and is responsible for many of the toxic effects associated with alcohol. The work of Dr Lieber (1995, 1997, 2000 and 2004) presents classic and scholarly reviews on the medical consequences of alcoholism, which are bedrock references for this literature overview (Chapter 2).

Alcohol, chemically known as ethanol ($\text{CH}_3\text{CH}_2\text{OH}$), is a colourless liquid containing a hydrophobic hydrocarbon end and a hydrophilic hydroxyl end, making it miscible in both aqueous and organic solutions. It can therefore easily diffuse across all cell membranes, which results in its being rapidly absorbed across the gastrointestinal tract and into the blood circulatory system, whence it is distributed throughout the body, exerting effects on most organ systems. For this reason, excessive alcohol consumption contributes to the development of many acute and chronic diseases.

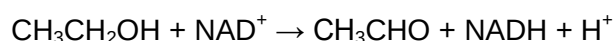
2.1 The metabolism of ethanol

Less than 10% of the ethanol absorbed by the body is eliminated through the kidneys and lungs. Most of the remaining 90% is metabolized mainly in the liver to acetaldehyde by three oxidative pathways. Each of these pathways depends on different enzymes and cofactors, and takes place within different subcellular compartments (Lieber, 1997) of the hepatocytes:

1. Cytoplasmic ethanol oxidation catalysed by the NAD^+ -dependent enzyme alcohol dehydrogenase (ADH).
2. Oxidation catalysed by microsomal enzymes of the endoplasmic reticulum, namely cytochrome P450 2E1 (CYP2E1), as well as (to a lesser extent) the CYP1A2 and CYP3A4 isoforms.
3. Oxidation catalysed by the peroxisomal enzyme catalase.

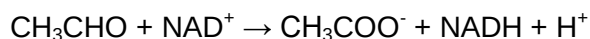
Pathways 2 and 3 are minor oxidative pathways of ethanol metabolism related mostly to chronic alcohol consumption (Ferrer-Dufol & Menao-Guillen, 2009; Lieber, 1995; Lieber, 1997; Lieber, 2000; Lieber, 2004; Zakhari & Li, 2007), and will not be discussed further.

During acute alcohol consumption, pathway 1 is responsible for approximately 90–95% of ethanol metabolism. This pathway involves the oxidation of ethanol to hydrogen and acetaldehyde (Manna *et al.*, 2010), in a reaction that is catalysed by ADH, an NAD^+ -dependent enzyme, and takes place in the cytoplasm of cells (Lieber, 1997). Various isoforms of this enzyme exist with varying affinities towards ethanol. ADH catalyses the reaction:



In this reaction ethanol loses a proton (H^+), and a hydride ion (H^-) is transferred from ethanol to NAD^+ to reduce it to NADH, resulting in an increased $\text{NADH}:\text{NAD}^+$ ratio (Das & Vasudevan, 2007; Lieber, 1997; Lieber, 2000; Lieber, 2004).

Acetaldehyde (produced in all three of the pathways listed above) is a highly reactive toxic compound that contributes to liver damage and is responsible for many of the toxic effects associated with ethanol consumption (Zakhari & Li, 2007). The acetaldehyde is oxidized to hydrogen and acetate in a reaction catalysed by the NAD^+ -dependent enzyme aldehyde dehydrogenase (ALDH), situated in the mitochondria of the cells (Lieber, 1997):



This reaction also involves the reduction of NAD^+ to NADH, further increasing the NADH pool and contributing to the increased $\text{NADH}:\text{NAD}^+$ ratio (Lieber, 2004).

Most of the acetate escapes from the liver into the blood circulation and is further metabolized to acetyl-CoA in a reaction catalysed by acetyl-CoA synthase 2 in cells of the heart, skeletal muscle, and brain, since, unlike liver cells, the mitochondria of these cells contain an abundance of this enzyme (Zakhari & Li, 2007). The acetyl-CoA can then participate in numerous metabolic pathways such as fatty acid biosynthesis (lipogenesis), amino acid metabolism, glycolysis, and the citric acid cycle (where it is metabolized to CO_2) (Garrett & Grisham, 2005; Manna *et al.*, 2010; Manna *et al.*, 2011).

Other pathways of ethanol elimination include:

1. A non-oxidative pathway, involving the enzyme fatty acid ethyl ester synthase, which plays a minor role in ethanol elimination, mainly in the heart and other organs that lack the capacity for oxidative ethanol metabolism (Lieber, 1997).
2. A very minor amount of ethanol (usually less than 0.1%) is converted to sulphate and glucuronide conjugates (ethyl sulphate and ethyl glucuronide) by the respective phase II detoxification enzymes in the liver, and are excreted in the urine. These molecules are good long-term markers of alcohol consumption, since their half-life in the urine is quite long and

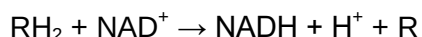
they remain detectable in serum, plasma, and hair for days after alcohol consumption (Bradford *et al.*, 2008; Wurst *et al.*, 2005).

3. Extrahepatic metabolism of ethanol, which is mostly limited to the stomach. Here, ethanol is oxidized by another isoform of ADH, which is activated by the extremely high gastric ethanol concentration that occurs after alcohol consumption. This gastric metabolism of ethanol (or first-pass metabolism) decreases the bioavailability of ethanol and represents a “protective barrier” against the systemic effects of alcohol consumption, at least when alcohol is consumed in small amounts, as in social drinking (Lieber, 1997). However, in the fasted state, ethanol is more rapidly passed from the stomach into the duodenum. This results in decreased first-pass metabolism when compared to the non-fasted state, and consequently a reduced “protective barrier”, increased blood alcohol concentrations, and more extensive alcohol-induced metabolic perturbations (DiPadova *et al.*, 1987).

2.2 Nicotinamide adenine dinucleotide (NAD)

The main oxidative pathway of ethanol metabolism involves two reactions catalysed by NAD⁺-dependent enzymes. This results in the reduction of NAD⁺ to NADH and the subsequent production of large amounts of NADH. Many of the effects of ethanol on metabolism result from this increased NADH production, and it is therefore of interest to understand what NAD is, how it functions and why it is so vital to many metabolic reactions.

NAD is a coenzyme composed of two nucleotides (ribose rings, one with adenine attached to the first carbon atom and the other with nicotinamide at this position), joined together by a bridge of two phosphate groups through the 5th carbon atoms. NAD is found in all living cells, and is involved in numerous vital metabolic enzymatic reactions, in which it serves as an electron carrier. NAD exists in two forms in cells: NAD⁺, which accepts electrons from other molecules, and NADH, which is used to donate electrons. During such an electron transfer (or redox) reaction (summarized in the formula below), two hydrogen atoms are removed from the substrate in the form of a hydride ion (H⁻) and a proton (H⁺). The proton is released into solution, the substrate (such as ethanol or acetaldehyde) is oxidized, and NAD⁺ is reduced to NADH by transfer of the hydride to the nicotinamide ring (Figure 2–1):



where RH₂ is the reduced form and R the oxidized form of the substrate.

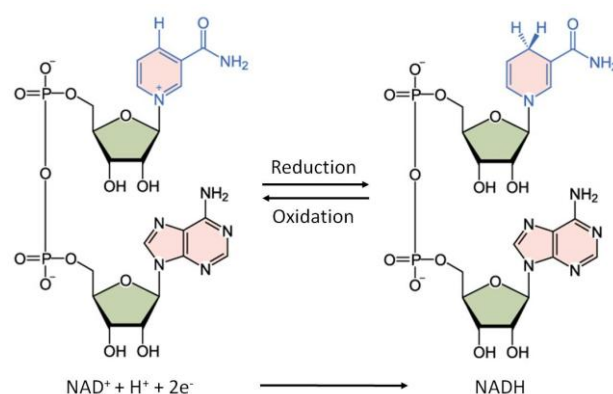


Figure 2-1: Schematic representation of the conversion between NAD⁺ and NADH during redox reactions [adapted from Rye *et al.*, 2016: “Download for free at <https://openstax.org/details/books/biology>”].

One particularly important area where redox reactions occur is in the release of energy from nutrients, such as glucose and fatty acids. During β -oxidation, glycolysis and the citric acid cycle, these compounds are oxidized, and their released energy is transferred to NAD⁺ by reduction to NADH. Since the mitochondrial membrane is impermeable to NADH, the electrons carried by the NADH produced in the cytoplasm are transferred into the mitochondria by mitochondrial shuttles such as the malate–aspartate shuttle, where they are used to reduce mitochondrial NAD⁺ to NADH. This mitochondrial NADH is then, again, oxidized to NAD⁺ by NADH dehydrogenase in the electron transport chain, where the electrons are transferred to coenzyme Q, and protons are pumped across a membrane to generate ATP through oxidative phosphorylation (Das & Vasudevan, 2007; Lin & Guarente, 2003; Zakhari & Li, 2007). Energy is released from ethanol in the same way. The large amount of NADH produced from ethanol oxidation after acute alcohol consumption however, overwhelms this system, leading to an increased NADH:NAD⁺ ratio, which is a hallmark of alcohol consumption (discussed in section 2.3 below) (Lieber, 2000). The increased NADH:NAD⁺ ratio in hepatocytes after acute alcohol consumption leads to alterations in several metabolic pathways, many of which seem to be attenuated with chronic alcohol consumption (Zakhari & Li, 2007).

The balance between the oxidized and reduced forms of NAD (or the NAD⁺:NADH ratio) plays an important role in regulating the intracellular redox state (a measurement that reflects both the metabolic activities and the health of cells), since the NAD⁺:NADH ratio controls the activity of several key metabolic enzymes (Lin & Guarente, 2003). It has been well documented that the NAD⁺:NADH ratio fluctuates in response to a change in metabolism, and many human diseases are associated with changes in this ratio (Lin & Guarente, 2003; Zakhari & Li, 2007). In healthy mammalian tissues, the cytoplasmic NAD⁺:NADH ratio is estimated to be about 600, favouring

oxidative reactions. The ratio of total $\text{NAD}^+:\text{NADH}$ is much lower, with estimates ranging from 3 to 10 (Lin & Guarente, 2003).

In addition to redox reactions, NAD is also involved in other cellular processes where NAD acts as a donor of ADP-ribose units, thereby destroying the NAD molecule (Khan *et al.*, 2007; Lin & Guarente, 2003). To replenish destroyed molecules, NAD can be synthesized *de novo* through the generation of quinolinic acid from the amino acid tryptophan (through the kynurenine pathway). Alcohol consumption, however, alters tryptophan metabolism and decreases the activity of tryptophan-2,3-dioxygenase, which impairs NAD biosynthesis. This further increases the $\text{NADH}:\text{NAD}^+$ ratio and exacerbates its effects (Manna *et al.*, 2011).

NAD is vital for proper cellular function and therefore the enzymes involved in NAD^+ -dependent reactions are often targets for drug discovery and research into future treatments for diseases (Khan *et al.*, 2007). Drug design and drug development exploit NAD in three ways: (1) as a direct target of drugs, (2) by designing enzyme inhibitors or activators, based on its structure, that change the activity of NAD^+ -dependent enzymes, and (3) by trying to inhibit NAD biosynthesis. The coenzyme NAD is not itself currently used as a treatment for any disease. NAD supplements are, however, available commercially and are recommended for their antioxidant and substance-abuse benefits, as well as for their ability to improve mental clarity, alertness and concentration, and to aid in the treatment of chronic fatigue.

2.3 The effect of acute alcohol consumption on metabolism

The US Department of Agriculture Dietary Guidelines define a drink as 360 mL of beer or spirit cooler, 150 mL of wine, or 45 mL of 80-proof distilled spirits. Such standard drinks contain approximately 12.5 g of ethanol (Zakhari & Li, 2007). Moderate (or low-level) drinking is defined as up to two drinks per day for men and up to one drink per day for women. Acute alcohol consumption (binge or “risky” drinking) is defined by the National Institute on Alcohol Abuse and Alcoholism as the consumption of five or more drinks by males or four or more drinks by females within two hours, resulting in a blood alcohol concentration of 0.08 g% or above (Roache *et al.*, 2015). Chronic or excessive alcohol consumption is defined as the consumption of more than five drinks by males or four drinks by females per day over a long period of time (Zakhari & Li, 2007).

Great variation in the pattern of drinking (quantity, frequency and duration) exists among alcohol consumers, and subsequently the health outcomes will be different for chronic alcohol drinkers compared to acute or moderate alcohol drinkers (Zakhari & Li, 2007). To date, most of the research done on the effects of alcohol consumption on human metabolism has involved

chronic alcohol consumption. However, many of these effects can also be seen after acute alcohol consumption, often only to a lesser extent.

Ethanol is a small molecule soluble in both water and lipids, and therefore permeates all tissues of the body and affects most vital functions (Lieber, 1997). Ethanol is involved in the perturbation of various metabolic pathways, altering the concentrations of metabolites in these pathways and leading to the production of unusual metabolites. Ethanol itself, its oxidation products, or these ethanol-induced altered concentrations and secondary metabolites can elicit a variety of pathogenic conditions (Gunzerath *et al.*, 2011). Since the liver is the primary organ for the oxidative metabolism of ethanol, it is the site for direct toxicity, and therefore many ethanol-induced disorders and diseases occur within the liver. Additionally, the lack of effective feedback control of the rate of hepatic ethanol metabolism may result in a displacement of up to 90% of the liver's normal metabolic substrates, which further explains why ethanol metabolism produces such striking metabolic imbalances in the liver.

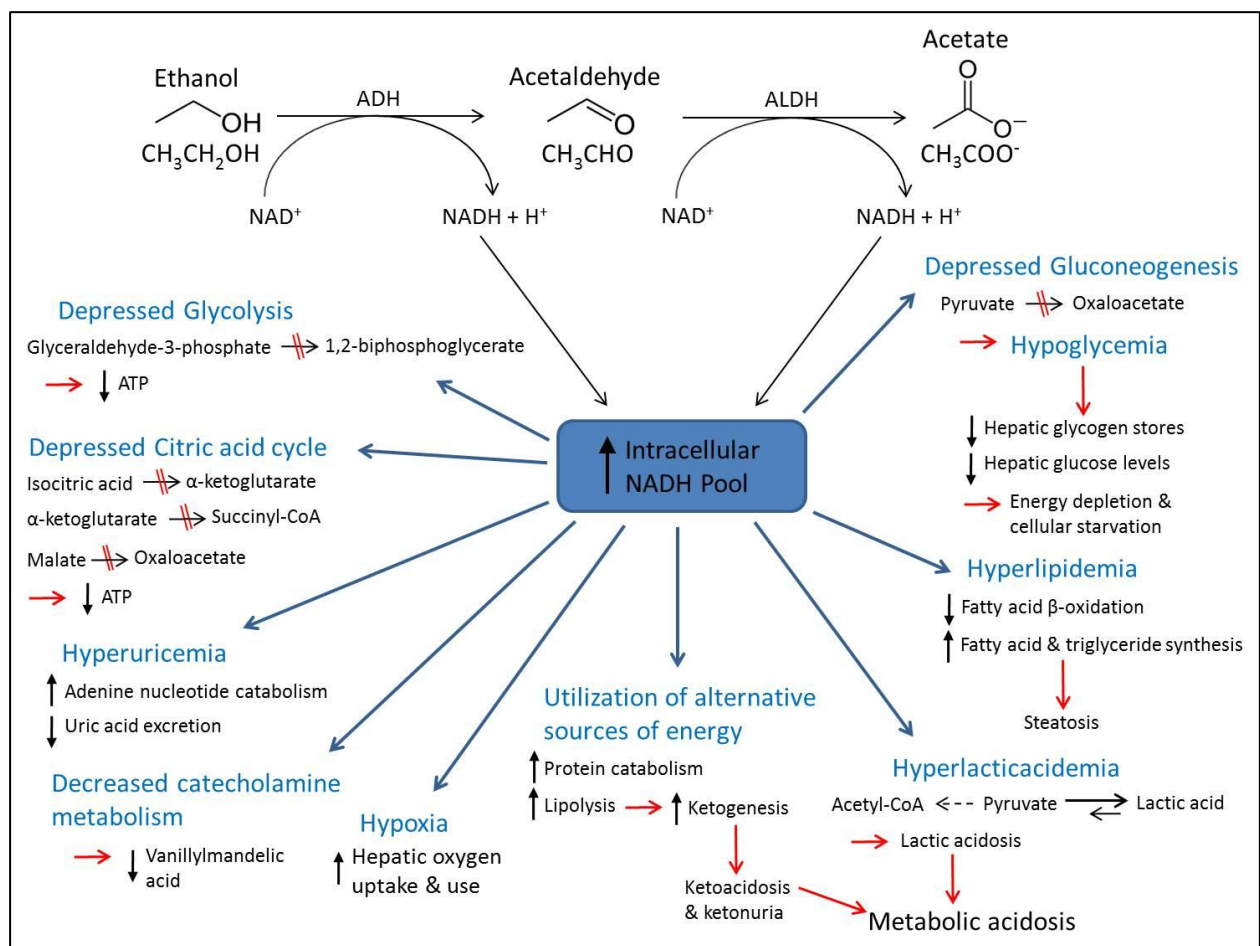


Figure 2-2: Summarized model of the global perturbations induced by excessive NADH (increased $\text{NADH}:\text{NAD}^+$ ratio) as a result of ethanol oxidation.

Most of the significant effects of ethanol on the metabolism result from the production of large amounts of NADH during the ADH- and ALDH-mediated oxidation of ethanol and acetaldehyde. This increases the NADH:NAD⁺ ratio in both the cytosol and mitochondria of hepatocytes, altering the redox state of the hepatocytes and inhibiting further energy production (Lieber, 1997). Several enzymes are regulated by the NADH:NAD⁺ ratio. Thus, the ethanol-induced increase of this ratio influences the directions of several reversible reactions. This results in widespread and noticeable global perturbations of numerous metabolic pathways within the human body, including alterations in hepatic lipid, carbohydrate, protein, lactic acid, and uric acid metabolism (illustrated in Figure 2–2) (Lieber, 1995; Zakhari & Li, 2007). NAD⁺-dependent dehydrogenase-catalysed reactions are particularly affected by the increased NADH:NAD⁺ ratio. These reactions are directly influenced by alcohol abuse and are therefore some of the primary sites of alcohol-induced perturbations. Inhibition of NAD⁺-dependent enzyme reactions due to the increased NADH:NAD⁺ ratio causes, amongst other effects:

- **Depressed glycolysis** due to the NADH-induced inhibition of the NAD⁺-dependent dehydrogenase sixth step of this pathway (conversion of glyceraldehyde-3-phosphate to 1,2-biphosphoglycerate). Thus, glucose is unable to be fully catabolized to pyruvate, leading to glucose no longer being a significant source of energy. The depression of this pathway also leads to the accumulation of certain glycolysis intermediates, such as glyceraldehyde-3 phosphate (Bradford & Rusyn, 2005).
- **Depressed citric acid cycle** due to saturation of NADH dehydrogenase, as well as feedback inhibition of the three NAD⁺-dependent dehydrogenase-catalysed reactions that occur within the citric acid cycle (Lieber, 1997; Lieber, 2000; Lieber, 2004). This results in the accumulation and increased excretion of various citric acid cycle intermediates, as well as a reduced capability to oxidize the excess acetyl-CoA (produced during ethanol oxidation) to excretable CO₂, which leads to the accumulation and increased excretion of acetate. Depression of the citric acid cycle also leads to the conversion of oxaloacetate to malate, leaving the oxaloacetate levels too low for citrate synthase to synthesize citrate for use in the citric acid cycle.
- Diminished ATP production within the glycolysis and citric acid cycle pathways, which lowers the energy status of the hepatocytes (Bradford & Rusyn, 2005).
- Decreased pyruvate dehydrogenase activity in the mitochondria, resulting in diminished conversion of pyruvate to acetyl-CoA (for use in the citric acid cycle), and subsequent increased conversion of pyruvate to lactic acid (increasing the lactic acid:pyruvate ratio). This reaction is catalysed by the NADH-dependent enzyme lactate dehydrogenase, and is favoured due to the presence of excess amounts of NADH. This increase in lactic acid

concentration results in **hyperlacticacidaemia** (increased blood lactic acid levels) and lactic acidosis (Lieber, 1997; Lieber, 2000; Lieber, 2004; Zakhari & Li, 2007).

- The accumulated glyceraldehyde-3-phosphate (due to depressed glycolysis) is metabolized by a NADH-mediated reaction to produce increased levels of glycerol-3-phosphate and decreased levels of inorganic phosphate. Inorganic phosphate acts as an inhibitor of AMP deaminase, a rate-limiting enzyme in the adenine nucleotide catabolism pathway. Decreased levels of inorganic phosphate thus remove the inhibition on AMP deaminase and result in increased adenine nucleotide catabolism. This causes the production of increased amounts of uric acid, which results in primary **hyperuricaemia** (increased blood uric acid levels) (Bradford & Rusyn, 2005; Garrett & Grisham, 2005; Lieber, 1997; Lieber, 2000; Mason, 2010; Witten *et al.*, 1973). In addition, lactic acid and ketone bodies (both of which are increased by increased ethanol metabolism) compete with uric acid for excretion in the distal tubules of the kidneys, resulting in reduced urinary excretion (and increased reabsorption) of uric acid, which leads to secondary hyperuricaemia (Lieber, 1997; Lieber, 2000). Hyperuricaemia provides an explanation for the observation of gout attacks following excessive alcohol consumption.
- **Reduced hepatic gluconeogenesis** (glucose production from pyruvate), due to decreased pyruvate concentrations and subsequent reduced pyruvate carboxylase activity (responsible for catalysing one of the rate-limiting steps of gluconeogenesis, namely the conversion of pyruvate to oxaloacetate). Pyruvate carboxylase is a biotin-dependent enzyme, activated by acetyl-CoA. Thus, its activity is further inhibited by insufficient biotin levels (due to ethanol-induced malnutrition and malabsorption) and low acetyl-CoA concentrations (due to decreased conversion of pyruvate to acetyl-CoA and the inability of hepatocytes to produce acetyl-CoA from the acetate produced during ethanol metabolism). Inhibition of the gluconeogenesis pathway results in hypoglycaemia (decreased blood glucose levels) (Lieber, 1997; Lieber, 2000).
- The alcohol-induced **hypoglycaemia** stimulates an adrenergic response (similar to the effects caused by epinephrine), which causes a rapid depletion of liver glycogen stores, an increase in fatty acid levels, and a massive, but temporary, increase in blood glucose levels (through the stimulation of glycolysis). Prolonged activation of glycolysis together with reduced hepatic gluconeogenesis and ethanol-induced depression of the citric acid cycle, results in decreased hepatic glucose levels, energy depletion and cellular starvation.
- Once the glycogen stores and blood glucose levels are depleted, and cellular starvation sets in, **pathways of alternative energy production** become activated. These include:

- Protein catabolism and use of amino acids as a source of energy, resulting in a decrease in hepatic protein content (Rachakonda *et al.*, 2014).
- Lipolysis (breakdown of lipids, due to increased lipolytic hormones), and subsequent increased ketogenesis (production of ketone bodies from fatty acids for energy), resulting in ketoacidosis (accumulation of ketone bodies in the blood) and ketonuria (excretion of ketone bodies in the urine). This, together with the hyperlacticacidaemia, leads to a state of massive **metabolic acidosis** (Kanetake *et al.*, 2005; Lefèvre *et al.*, 1970; Zakhari & Li, 2007). Ketogenesis is also stimulated by the high concentration of circulating acetyl-CoA, which further enhances the ketoacidosis. β -Hydroxybutyrate is the primarily formed ketone body, due to the NADH-dependent conversion of acetoacetic acid to β -hydroxybutyrate.
- **Depressed fatty acid β -oxidation**, since the mitochondria will preferentially use the hydrogen equivalents originating from ethanol oxidation, rather than those derived from the oxidation of fatty acids that normally serve as the main energy source of the liver (Lieber, 2000). Concurrently, increased fatty acid and triglyceride synthesis occurs due to accumulation of acetyl-CoA, decreased hepatic release of lipoproteins, increased mobilization of peripheral fat, enhanced hepatic uptake of circulating lipids, and decreased fatty acid β -oxidation (Lieber, 1997; Lieber, 2000). This causes **hyperlipidaemia** (increased blood lipid levels) and hypertriglyceridaemia (increased blood triglyceride levels), which leads to accumulation and deposition of fat in the liver (Lieber, 2004). The increased glycerol-3-phosphate concentration also favours hepatic triglyceride accumulation by trapping fatty acids, which contributes to the ethanol-induced hyperlipidaemia (Lieber, 1997; Lieber, 2000; Zakhari & Li, 2007).
- Both short- and long-term alcohol consumption is known to cause acceleration of the rates of ethanol metabolism, and increased oxygen uptake by hepatocytes (Lieber, 2000). Ethanol directly increases the uptake of molecular oxygen by inducing the CYP2E1 pathway (which requires molecular oxygen for the catalytic activation of CYP2E1), and also by the NADH-induced increase in the activity of the mitochondrial electron transport chain. Ethanol indirectly increases oxygen use through lipopolysaccharide-induced activation of Kupffer cells, as well as through stimulation of general hepatocyte metabolic activity. The increased oxygen uptake results in significant **hypoxia** (insufficient oxygen) in the perivenous hepatocytes, which activates glycolysis and causes increased production and excretion of lactic acid, which in turn contributes to lactic acidosis (Bradford *et al.*, 2008; Zakhari & Li, 2007).

- **Decreased catabolism of epinephrine and norepinephrine** due to the inhibition of the NAD⁺-dependent ALDH-catalysed reactions in these pathways. However, catecholamine (especially norepinephrine) metabolism is also altered due to competition for ADH and ALDH between catecholamine intermediates and ethanol and acetaldehyde, respectively: (1) Acute alcohol consumption results in altered norepinephrine metabolism, since class I ADH enzymes also catalyse the oxidation of the intermediary alcohols 4-hydroxy-3-methoxyphenyl glycol (HMPG) and 3,4-dihydroxyphenyl glycol during norepinephrine catabolism. This altered metabolism is indicated by increased excretion of HMPG as well as decreased excretion of vanillylmandelic acid, the downstream product of HMPG (Mardh *et al.*, 1985). (2) The presence of excess acetaldehyde has also been shown to inhibit catecholamine degradation (Witten *et al.*, 1973).

2.4 The effect of chronic alcohol consumption on metabolism

Alcoholism is a common problem worldwide. Chronic alcohol consumption and abuse is detrimental to overall health and remains one of the main factors that cause liver injury and affect liver function. This can result in a number of serious pathological and metabolic disturbances and diseases. Alcoholic liver disease (ALD) is a common complication of alcohol misuse, and a major cause of alcohol-related morbidity and mortality worldwide (Li *et al.*, 2014). ALD is characterized by several distinct stages of liver injury ranging from steatosis to cirrhosis, and is mainly caused by the alcohol-induced change in the redox state of hepatocytes, reactive oxygen species (ROS), and the toxicity of acetaldehyde (Zakhari & Li, 2007).

Many of the effects of chronic alcohol consumption on metabolism are only more severe versions of the effects observed after acute alcohol consumption. However, there are various other pathophysiological conditions associated only with chronic alcohol consumption:

- Ethanol is a substantial source of metabolic energy (providing approximately 29.7 kJ per gram, a value that exceeds the energy content of both carbohydrates and proteins), and accounts on average for half of an alcoholic's caloric intake. As such, normal nutrients are often replaced, which results in **malnutrition** and deficiencies of various essential vitamins. Additionally, alcohol-induced mucosal damage to the gastrointestinal tract causes inhibition of digestion, absorption and secretion of nutrients, resulting in secondary malnutrition (Lieber, 1995; Lieber, 1997; Lieber, 2004; Rajendram & Preedy, 2005; Zivkovic, 2009).
- Vitamin B1 (thiamine) is essential for the optimal functioning and health of the cardiovascular, nervous, and digestive systems, and is also essential for the conversion of carbohydrates into energy (since it is a co-factor required by three enzymes that catalyse various steps in the breakdown of glucose). Alcohol impairs thiamine absorption and

utilization, which results in **thiamine deficiency** and contributes to the neuro-degeneration often seen in alcoholics (Harrigan *et al.*, 2008).

- As in acute alcohol consumption, **ketoacidosis** and **ketonuria** are caused by chronic alcohol consumption. An increased β -hydroxybutyrate:acetoacetic acid ratio is indicative of ethanol-induced ketoacidosis, and it has been suggested that increased β -hydroxybutyrate levels could be a marker of unexpected death in alcoholics due to hypoglycaemia (Kanetake *et al.*, 2005; Lefèvre *et al.*, 1970).
- ALDH activity is significantly reduced by chronic alcohol consumption, which results in a decreased capability of mitochondria to oxidize acetaldehyde. This, together with unaltered, or even enhanced, rates of ethanol oxidation (due to induction of CYP2E1, and therefore increased acetaldehyde production) results in an **accumulation of acetaldehyde** (Lieber, 1997). Acetaldehyde itself has many toxic effects (which lead to the development of hepatocyte toxicity), including (Lieber, 1995; Lieber, 1997; Lieber, 2000; Lieber, 2004):
 - Depleting reduced glutathione (GSH) levels, which increases the toxic effect of free radicals, and results in mitochondrial damage and the promotion of cell death.
 - Markedly reducing oxygen utilization in hepatic mitochondria damaged by long-term alcohol consumption.
 - Formation of highly unstable protein and DNA adducts, which may inactivate enzymes, alter DNA repair mechanisms, stimulate fibrogenesis and collagen accumulation, and induce immune responses (Das & Vasudevan, 2007; Lieber, 1997; Lieber, 2000).
 - Binding to the tubulin of microtubules, thereby blocking the secretion of proteins, and leading to the accumulation of proteins, lipids, water, and electrolytes in hepatocytes. This causes hepatocytes to enlarge, or “balloon”, which is a hallmark of ALD (Lieber, 2000; Lieber, 2004).
 - Crossing the placenta, impairing fetal DNA methylation, and contributing to fetal alcohol syndrome.

However, the acute reaction due to acetaldehyde accumulation (causing nausea, headache, vomiting, and possibly convulsions), which may be enhanced by competitive inhibition of ALDH by certain drugs (such as disulfiram) when taken with alcohol, may have possible clinical applications in the treatment of some alcoholic patients (Lieber, 1997).

- It is well-established that the prevalence of **hypertension** is higher among alcohol-dependent patients than among the normal population. A possible explanation for this is the accumulation of succinic acid and α -ketoglutaric acid (due to the alcohol-induced reduction of citric acid cycle activity), which have been found to increase blood pressure (Harrigan *et al.*, 2008). The vasopressor effect (resulting in the contraction of blood vessels) of ethanol

may also explain the association between long-term alcohol consumption and hypertension (Lieber, 1995).

- **Mitochondrial damage** in the hepatocytes of alcoholics due to the toxic effects of ethanol and acetaldehyde on these organelles, as well as the result of ROS-mediated oxidative stress (Lieber, 2004). This damage could be responsible for the reduced muscle strength and impaired central nervous system function frequently found in alcoholics (Mancinelli, 2014). Mitochondrial damage may also promote oxidative stress and altered lipid metabolism (which contributes to liver damage), and may also be associated with hypertension (Harrigan *et al.*, 2008; Lieber, 2004).
- Excessive alcohol consumption is associated with the increased risk for and incidence of various **cardiovascular disorders**, including atherosclerosis and fibrosis, cardiac myopathy, myocardial hypertrophy, arrhythmias, and haemorrhagic and ischaemic strokes (Hines & Rimm, 2001; Lieber, 1995; Numminen *et al.*, 2000; Steiner *et al.*, 2015; Zheng *et al.*, 2015). This is in contrast to light or moderate alcohol consumption, which has been shown to be beneficial for the cardiovascular system, and is associated with a lower risk of coronary heart disease, peripheral arterial disease and ischaemic strokes (Hines & Rimm, 2001; Lieber, 1995; Numminen *et al.*, 2000).
- As with acute alcohol consumption, excessive alcohol consumption is known to cause alcohol-induced **alterations to lipid metabolism**, such as increased lipogenesis (lipid production from the excess acetyl-CoA generated during the metabolism of ethanol), increased cholesterol synthesis and secretion, decreased β -oxidation, enhanced microsomal ω -oxidation and peroxisomal β -oxidation of fatty acids (to compensate partly for the decreased hepatic β -oxidation), and dysfunctional phospholipid synthesis (with subsequent decreased plasma membrane remodelling). These changes result in **hyperlipidaemia** and **hypercholesterolaemia** (increased blood cholesterol levels), both of which are risk factors for coronary heart disease. This might partially explain the enhanced incidence of cardiovascular disorders and fatal myocardial infarctions associated with alcoholism (Das & Vasudevan, 2007; Lieber, 1997; Lieber, 2004; Visioli *et al.*, 1998).
- The altered lipid metabolism also causes lipid accumulation and deposition in the liver, eventually causing **steatosis** (fatty liver) (Li *et al.*, 2014; Lieber, 2000; Lieber, 2004; Manna *et al.*, 2010). If left untreated, steatosis can progress to **alcoholic steatohepatitis** (alcohol-induced liver inflammation together with hepatic lipid accumulation), a systemic condition that affects 10–35% of individuals with alcoholism (Zivkovic *et al.*, 2009). Alcoholic steatohepatitis causes further complications including **fibrosis** and **cirrhosis** (excessive formation of fibrous connective tissue in the liver, leading to scarring, hepatocyte damage,

hepatomegaly, distorted hepatic morphology and jaundice), and is one of the main causes of end-stage liver disease worldwide (Bradford *et al.*, 2008; Lieber, 2000; Lieber, 2004; Petts *et al.*, 2014; Steiner *et al.*, 2015; Zakhari & Li, 2007; Zivkovic *et al.*, 2009).

- Alcohol-induced mucosal damage to the gastrointestinal tract compromises the gastrointestinal immune system, resulting in increased susceptibility to infections (Rajendram & Preedy, 2005), which, together with the acetaldehyde-induced hepatotoxicity, also contributes to the development of **alcoholic hepatitis** (alcohol-induced necrosis and inflammation of the liver) (Lieber, 2000).
- Chronic alcohol consumption and subsequent cirrhosis increase the risk of the development of liver cancer (**hepatocellular carcinoma**), one of the leading cancers worldwide in terms of prevalence and mortality (Nahon *et al.*, 2012). Alcohol abuse is also associated with cancers of the alimentary and respiratory tracts, as well as possibly with pancreatic, gastric and breast cancer (Garro & Lieber, 1990).
- Chronic alcohol consumption results in increased CYP2E1 activity, which contributes to the accelerated hepatic metabolism of ethanol and the **metabolic tolerance to ethanol** observed in alcoholics, which results in the consumption of higher doses of alcohol. CYP2E1 induction also accelerates the metabolism and elimination of other drugs, resulting in increased **drug tolerance** (Lieber, 1997; Lieber, 2000; Lieber, 2004). However, the presence of high alcohol concentrations may also cause **inhibition of drug metabolism**. This is mainly due to the competition of certain drugs with ethanol for detoxification by CYP2E1 (Ferrer-Dufol & Menao-Guillen, 2009; Lieber, 1995; Lieber, 1997), but also due to the NADH-induced depression of the citric acid cycle, which could cause depletion of intermediates necessary for the generation of cytosolic NADPH (a key substrate for CYP2E1 activity). These inhibitory effects may explain the observation that simultaneous administration of ethanol and drugs slows the rate of drug metabolism (Lieber, 1997).
- Ethanol-induced increased CYP2E1 activity also results in the increased activation of various other xenobiotic compounds into more polar toxic forms, thereby increasing alcoholics' **vulnerability to these xenobiotics** (Lieber, 1997; Lieber, 2000).
- Chronic alcohol consumption and subsequent increased CYP2E1 activity greatly **enhances ROS production** (Lieber, 2004). Alcohol can also alter the levels of certain metals in the body, thereby facilitating ROS production. High levels of ROS are toxic to cells because they react with cellular macromolecules, denature proteins (inactivate enzymes), and increase the risk of DNA- and tissue damage, and cell death (Das & Vasudevan, 2007; Kim *et al.*, 2008; Ojeda *et al.*, 2015; Wu & Cederbaum, 2005). Antioxidant pathways are activated to

help protect the body when challenged by chronic exposure to xenobiotics. GSH is a critical cellular antioxidant, which limits the toxicity of ethanol and other toxic chemicals by neutralizing ROS (Lieber, 1997; Lieber, 2000; Lieber, 2004). However, the high levels of ROS produced during chronic alcohol consumption results in GSH depletion, which impairs the antioxidant power of the cells, resulting in alcohol-induced **oxidative stress** (Lieber, 2004). Ethanol itself also reduces the activities of antioxidants and antioxidative systems that can eliminate ROS, and exacerbates GSH depletion by inhibiting the mitochondrial glutathione transport system (Das & Vasudevan, 2007). In patients with cirrhosis, hepatic depletion of α -tocopherol, a major antioxidant, further increases oxidative stress (Lieber, 1995).

- **Oxidative stress** is associated with mitochondrial damage, increased lipid and protein oxidation, alcohol-induced liver injury, and DNA mutations, alteration and epigenetic modifications (since DNA is very sensitive to the oxidative stress effect). Oxidative stress also increases the risk for various pathogenic conditions, including, atherosclerosis, radiation damage, ischaemia reperfusion injury, diabetes mellitus, neurodegenerative diseases, hypoxia, toxic injuries (including acute and chronic alcohol toxicity), carcinogenesis and the release of inflammatory cytokines, and the development of ALD (including hepatic inflammation, liver damage, fibrosis and necrosis) (Bradford & Rusyn, 2005; Das & Vasudevan, 2007; Li *et al.*, 2011; Lieber, 2000; Lieber, 2004).

Ethanol-induced ROS production, antioxidant depletion and increased oxidative stress can potentially be counteracted through: (1) pretreatment with hepatocyte growth factor (which increases the GSH:GSSG (oxidized glutathione) ratio and enhances the antioxidant activity of the cells (Valdés-Arzate *et al.*, 2009)); (2) generation of NADPH through the pentose phosphate pathway (Rachakonda *et al.*, 2014); (3) ethanol-induced hyperuricaemia (since uric acid acts as a free radical scavenger) (Mason, 2010); (4) supplementation with S-adenosyl methionine (SAME) (which can be used to replenish GSH) (Lieber, 2000; Lieber, 2004); and (5) supplementation with polyenylphosphatidylcholine (which repletes hepatic phospholipids, thereby preventing membrane abnormalities and fibrosis) (Lieber, 1997; Lieber, 2000; Lieber, 2004).

In conclusion: Following this literature overview, it is apparent that a holistic perspective on the effect of acute alcohol consumption seems to be lacking. It is anticipated that this metabolomics investigation of acute alcohol consumption would:

1. Provide a holistic view of the comprehensive and complex metabolic consequences due to alcohol consumption discussed above.
2. Reveal insights not observed through the traditional biochemical investigations.

REFERENCES

- Bradford, B.U. & Rusyn, I. 2005. Swift increase in alcohol metabolism (SIAM): understanding the phenomenon of hypermetabolism in liver. *Alcohol*, 35(1):13–17.
- Bradford, B.U., O'Connell, T.M., Han, J., Kosyk, O., Shymonyak, S., Ross, P.K., Winnike, J., Kono, H. & Rosyn, I. 2008. Metabolomic profiling of a modified alcohol liquid diet model for liver injury in the mouse uncovers new markers of disease. *Toxicology and applied pharmacology*, 232:236–243.
- Das, S.K. & Vasudevan, D.M. 2007. Alcohol-induced oxidative stress. *Life sciences*, 81:177–187.
- DiPadova, C., Worner, T.M., Julkunen, R.J. & Lieber, C.S. 1987. Effects of fasting and chronic alcohol consumption on the first-pass metabolism of ethanol. *Gastroenterology*, 92(5):1169–1173.
- Ferrer-Dufol, A. & Menao-Guillen, S. 2009. Toxicogenomics and clinical toxicology: an example of the connection between basic and applied sciences. *Toxicology letters*, 186(1):2–8.
- Garrett, R.H. & Grisham, C.M. 2005. Biochemistry. 3rd ed. Belmont, CA: Thomson Brooks/Cole.
- Garro, A.J. & Lieber, C.S. 1990. Alcohol and cancer. *Annual review of pharmacology and toxicology*, 30(1):219–249.
- Gunzerath, L., Hewitt, B.G., Li, T-K. & Warren, K.R. 2011. Alcohol research: past, present, and future. *Annals of the New York Academy of Sciences*, 1216(1):1–23.
- Harrigan, G.G., Maguire, G. & Boros, L. 2008. Metabolomics in alcohol research and drug development. *Alcohol research & health*, 31(1):26–35.
- Hines, L.M. & Rimm, E.B. 2001. Moderate alcohol consumption and coronary heart disease: a review. *Postgraduate medical journal*, 77:747–752.
- Kanetake, J., Kanawaku, Y., Mimasaka, S., Sakai, J., Hashiyada, M., Nata, M. & Funayama, M. 2005. The relationship of a high level of serum beta-hydroxybutyrate to cause of death. *Legal medicine*, 7(3):169–174.

- Khan, J.A., Forouhar, F., Tao, X. & Tong, L. 2007. Nicotinamide adenine dinucleotide metabolism as an attractive target for drug discovery. *Expert opinion on therapeutic targets*, 11(5):695–705.
- Kim, S.J., Jung, Y.S., Kwon, D.Y. & Kim, Y.C. 2008. Alleviation of acute ethanol-induced liver injury and impaired metabolomics of S-containing substances by betaine supplementation. *Biochemical and biophysical research communications*, 368(4):893–898.
- Lefèvre, A., Adler, H. & Lieber, C.S. 1970. Effect of ethanol on ketone metabolism. *Journal of clinical investigation*, 49(10):1775–1782.
- Li, H-H., Tyburski, J.B., Wang, Y-W., Strawn, S., Moon, B-H., Kallakury, B.V.S., Gonzalez, F.J. & Fornace Jr., A.J. 2014. Modulation of fatty acid and bile acid metabolism by peroxisome proliferator-activated receptor α protects against alcoholic liver disease. *Alcoholism: clinical and experimental research*, 38(6):1520–1531.
- Li, S., Liu, H., Jin, Y., Lind, S., Caic, Z. & Jiang, Y. 2011. Metabolomics study of alcohol-induced liver injury and hepatocellular carcinoma xenografts in mice. *Journal of chromatography B*, 879(24):2369–2375.
- Lieber, C.S. 1995. Medical disorders of alcoholism. *New England journal of medicine*, 333(16):1058–1065.
- Lieber, C.S. 1997. Ethanol metabolism, cirrhosis and alcoholism. *Clinica chimica acta*, 257(1):59–84.
- Lieber, C.S. 2000. Alcoholic liver disease: new insights in pathogenesis lead to new treatments. *Journal of hepatology*, 32:113–128.
- Lieber, C.S. 2004. Alcoholic fatty liver: its pathogenesis and mechanism of progression to inflammation and fibrosis. *Alcohol*, 34(1):9–19.
- Lin, S-J. & Guarente, L. 2003. Nicotinamide adenine dinucleotide, a metabolic regulator of transcription, longevity and disease. *Current opinion in cell biology*, 15(2):241–246.
- Mancinelli, R. 2014. Fetal alcohol spectrum disorders: can alcohol-related oxidative stress concur to the prenatal damage? *OA alcohol*, 2(1):9.

Manna, S.K., Patterson, A.D., Yang, Q., Krausz, K.W., Li, H., Ilde, J.R., Fornace Jr., A.J. & Gonzalez, F.J. 2010. Identification of noninvasive biomarkers for alcohol-induced liver disease using urinary metabolomics and the *Ppara*-null mouse. *Journal of proteome research*, 9(8):4176–4188.

Manna, S.K., Patterson, A.D., Yang, Q., Krausz, K.W., Ilde, J.R., Fornace Jr., A.J. & Gonzalez, F.J. 2011. UPLC–MS-based urine metabolomics reveals indole-3-lactic acid and phenyllactic acid as conserved biomarkers for alcohol-induced liver disease in the *Ppara*-null mouse model. *Journal of proteome research*, 10(9):4120–4133.

Mardh, G., Luehr, C.A. & Vallee, B.L. 1985. Human class I alcohol dehydrogenases catalyze the oxidation of glycols in the metabolism of norepinephrine. *Proceedings of the National Academy of Sciences USA*, 82(15):4979–4982.

Mason, S.W. 2010. The metabolomics of acute alcohol abuse. Potchefstroom: NWU. (Dissertation – M.Sc.).

Nahon, P., Amathieu, R., Triba, M.N., Bouchemal, N., Nault, J-C., Zioli, M., Seror, O., Dhonneur, G., Trinchet, J-C., Beaugrand, M. & Le Moyec, L. 2012. Identification of serum proton NMR metabolomic fingerprints associated with hepatocellular carcinoma in patients with alcoholic cirrhosis. *Clinical cancer research*, 18(24):6714–6722.

Numminen, H., Syrjälä, M., Benthin, G., Kaste, M. & Hillborn, M. 2000. The effect of acute ingestion of a large dose of alcohol on the hemostatic system and its circadian variation. *Stroke*, 31(6):1269–1273.

Ojeda, M.L., Rua, R.M., Murillo, M.L., Carreras, O. & Nogales, F. 2015. Binge drinking during adolescence disrupts Se homeostasis and its main hepatic selenoprotein expression. *Alcoholism: clinical and experimental research*, 39(5):818–826.

Petts, G., Lloyd, K. & Goldin, R. 2014. Fatty liver disease. *Diagnostic histopathology*, 20(3):102–108.

Rachakonda, V., Gabbert, C., Raina, A., Bell, L.N., Cooper, S., Malik, S. & Behari, J. 2014. Serum metabolomic profiling in acute alcoholic hepatitis identifies multiple dysregulated pathways. *PLOS ONE*, 9(12):e113860. doi:10.1371/journal.pone.0113860.

- Rajendram, R. & Preedy, V.R. 2005. Effect of alcohol consumption on the gut. *Digestive diseases*, 23(3–4):214–221.
- Roache, J.D., Karns, T.E., Hill-Kapturczak, N., Mullen, J., Liang, Y., Lamb, R.J. & Dougherty, D.M. 2015. Using transdermal alcohol monitoring to detect low-level drinking. *Alcoholism: clinical and experimental research*, 39(7):1120–1127.
- Rye, C., Wise, R., Jurukovski, V., DeSaix, J., Choi, J. & Avissar, Y. 2016. Biology. Houston, TX: OpenStax Rice University. <https://openstax.org/details/books/biology> in your citation. Date of access: 4 Nov. 2017.
- Steiner, J.L., Pruznak, A.M., Navaratnarajah, M. & Lang, C.H. 2015. Alcohol differentially alters extracellular matrix and adhesion molecule expression in skeletal muscle and heart. *Alcoholism: clinical and experimental research*, 39(8):1330–1340.
- Valdés-Arzate, A., Luna, A., Bucio, L., Licona, C., Clemens, D.L., Souza, V., Hernandez, E., Kershenovich, D., Gutiérrez-Ruiz, M.C. & Gómez-Quiroz, L.E. 2009. Hepatocyte growth factor protects hepatocytes against oxidative injury induced by ethanol metabolism. *Free radical biology & medicine*, 47(4):424–430.
- Visioli, F., Monti, S., Colombo, C. & Galli, C. 1998. Ethanol enhances cholesterol synthesis and secretion in human hepatoma cells. *Alcohol*, 15(4):299–303.
- Witten, T.A., Levine, S.P., King, J.O. & Markey, S.P. 1973. Gas-chromatographic–mass-spectrometric determination of urinary acid profiles of normal young adults on a controlled diet. *Clinical chemistry*, 19(6):586–589.
- Wu, D. & Cederbaum, A.I. 2005. Oxidative stress mediated toxicity exerted by ethanol-inducible CYP2E1. *Toxicology and applied pharmacology*, 207(2):570–576.
- Wurst, F.M., Alling, C., Aradottir, S., Pragst, F., Allen, J.P., Weinmann, W., Marmillot, P., Ghosh, P., Lakshman, R., Skipper, G.E., Neumann, T., Spies, C., Javors, M., Johnson, B.A., Ait-Daoud, N., Akhtar, F., Roache, J.D. & Litten, R. 2005. Emerging biomarkers: new directions and clinical applications. *Alcoholism: clinical and experimental research*, 29(3):465–473.
- Zakhari, S. & Li, T-K. 2007. Determinants of alcohol use and abuse: impact of quantity and frequency patterns on liver disease. *Hepatology*, 46(6):2032–2039.

Zivkovic, A.M., German, J.B., Esfandiari, F. & Halsted, C.H. 2009. Quantitative lipid metabolomic changes in alcoholic micropigs with fatty liver disease. *Alcoholism: clinical and experimental research*, 33(4):751–758.

Zheng, L., Wu, X., Dong, X., Ding, X. & Song, C. 2015. Effects of chronic alcohol exposure on the modulation of ischemia-induced glutamate release via cannabinoid receptors in the dorsal hippocampus. *Alcoholism: clinical and experimental research*, 39(10):1908–1916.

PART II NMR METABOLOMICS OF ACUTE ALCOHOL CONSUMPTION

CHAPTER 3 METABOLOMICS AND ALCOHOLISM

3.1 What is metabolomics?

Many scholarly reviews have been published on metabolomics (Fiehn *et al.*, 2007; Goodacre *et al.*, 2004; Goodacre *et al.*, 2007), metabolomics technology (Koek *et al.*, 2011; Weljie *et al.*, 2006; Zhou *et al.*, 2012), and its applications (Bundy *et al.*, 2009; Spratlin *et al.*, 2009; Zhang *et al.*, 2012). It is therefore beyond the purpose of my thesis to contribute to a deeper understanding of metabolomics and its technologies. However, a brief and broad outline of metabolomics is presented here in order to justify the use of such an approach in my study.

Metabolomics is fast becoming an established field in which comprehensive analytical methods are used to study metabolism, metabolites and other small molecules within cells and tissues in order to better understand the global metabolic changes in biological systems (Castle *et al.*, 2006). The aim of metabolomics is to obtain new insights into biochemical processes and pathways by studying relationships between metabolites (that is, low-molecular-weight compounds ($M_w < 1500$ Da), including lipids, sugars, amino acids, nucleotides, and several different organic molecules, that serve as substrates and/or products in various metabolic pathways) (Bernini *et al.*, 2011). An advantage of metabolomics is that the analyst can simultaneously analyse metabolites originating from multiple metabolic pathways, thereby providing a global analysis of the perturbed metabolites, and capturing the status of diverse biochemical pathways at a particular moment in time. The term metabolomics, first used by Oliver Fiehn in 2001, was thus appropriately defined as "a comprehensive analysis in which all metabolites of a biological system were identified and quantified" (Fiehn, 2001). However, to date, no analytical method exists that is capable of comprehensively and accurately measuring all metabolites in a biological system simultaneously, mostly because of the chemical complexity and heterogeneity of metabolites, the extraction protocols, and the dynamic concentration range that needs to be measured (Goodacre *et al.*, 2004). Therefore, metabolomics is commonly described as the study of the metabolic profiles of organisms and the accurate identification and quantification of the concentration changes of many metabolites in these profiles that occur due to either natural fluctuations or due to any endogenous or exogenous perturbations (Hendriks *et al.*, 2011; Manna *et al.*, 2010), such as disease, therapeutic intervention, genetic modification, environmental variation or, as in the case of the study reported here, ethanol consumption. Variables identified in metabolomics analyses reflect

the biochemical pathways from which they originate (Harrigan *et al.*, 2008), and significant changes in the metabolite profiles seen in group comparisons (for example, males vs females; sick vs healthy individuals, etc.) are thus indicators of perturbations in these pathways.

As in many other research fields, the biological questions addressed in metabolomics investigations are diverse. However, despite this, most metabolomics approaches follow a similar pattern or workflow — starting with the biological research question, encompassing the experimental design and methods of data collection and analysis, and ending with a biological interpretation of the results and the gaining of new insights (Dunn *et al.*, 2008; Fiehn *et al.*, 2007; Goodacre *et al.*, 2007; Hendriks *et al.*, 2011). Each step in this workflow can potentially give rise to experimental variation (Chapter 5), and, since metabolomics is such a complicated interdisciplinary research field, close collaboration between researchers in different fields — including analytical chemistry, organic chemistry, chemometrics, informatics and bioscience — is required to construct a robust system in which these variations are minimized (Fukusaki & Kobayashi, 2005). An example of a typical metabolomics workflow is illustrated in Figure 3–1.

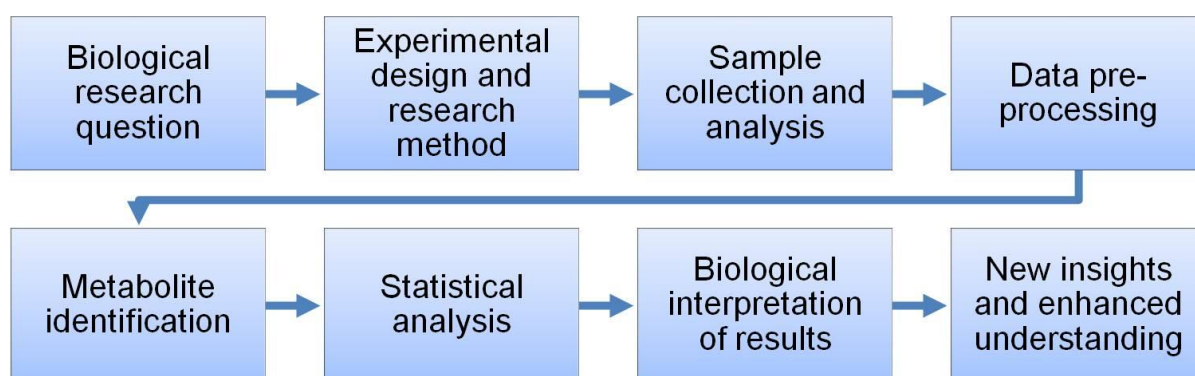


Figure 3-1: Metabolomics workflow [adapted with permission from Hendriks *et al.*, 2011].

Metabolomics studies can either follow an “open mind” approach with the aim of gaining new information relating to a proposed research question, or else certain methods and techniques can be used in order to test a previously formulated hypothesis (Dunn *et al.*, 2008). Although the metabolism of ethanol has been extensively studied (Chapter 2), the two main outcomes envisaged for my thesis (“a holistic view on acute alcohol consumption” and “insights not previously observed through traditional biochemical investigations”) imply that I followed an “open mind” approach. Consequently, the initial observations in all experimental sections were exploratory in nature. However, the data analysis provided new quantitative information in all the sections which could be disseminated through peer-reviewed publications.

The method of data generation in metabolomics is highly dependent on the aim of the study. Data generation can broadly be characterized as either untargeted or targeted:

- Untargeted metabolomics studies may be applied in hypothesis-testing investigations, but generally follow an “open mind” approach, which often result in the formulation of new hypotheses (Kristensen *et al.*, 2012). Untargeted investigations seek to assign as many individual chromatographic or spectroscopic peaks as possible (Vinayavekhin & Saghatelian, 2010) in order to provide a global overview of the readily detectable metabolites within a sample (Harrigan *et al.*, 2008). Untargeted metabolomics studies enable the identification of the broad metabolic profiles of samples, and produce a comprehensive list of the metabolites within a sample, with either absolute or relative quantification. These lists can then be used to find novel metabolites that can be used as biomarkers of disease or other conditions under investigation. The main advantage of untargeted studies is that these investigations give a general overview of metabolic variability, rather than focusing on a preselected set of target metabolites. However, these types of investigations are generally tedious to apply to large numbers of biological samples, since they often include time-consuming analytical and quantification steps, such as identification.
- In targeted metabolomic studies, the focus is placed on a certain number of pre-determined metabolites, group of metabolites, or metabolite class (amino acids, organic acids, lipids, etc.). In these studies, the chemical compounds are first separated, typically through gas chromatography (GC) or liquid chromatography (LC). This is followed by metabolite annotation through comparison of the spectra of interest with a library of reference spectra of pure compounds — the separated compounds are characterized by name, through annotation from the library information. The annotated metabolites are then quantified, followed by selection of the important metabolites, based on unbiased statistical analysis and inclusion and exclusion criteria, where applicable, in order to address certain specific biological questions formulated for the study (Robertson *et al.*, 2011; Steinfath *et al.*, 2008).

Targeted studies, where a limited number of metabolites are studied, are mostly quantitative, whereas untargeted studies are often qualitative and only semi-quantitative, since the focus of these studies is rather placed on the identification of metabolites and the determination of metabolite profiles (Robertson *et al.*, 2011). Because the exact identity of the analytes in targeted studies is known, it is sometimes possible to measure their absolute concentrations with the use of appropriate standards — a specific amount of a pure standard (a stable isotope of a specific metabolite, for example) of each of the targeted metabolites can be added to the biofluid sample analysed, so that the concentrations of the targeted metabolites can be determined accurately. Due to the large number of metabolites involved in untargeted studies,

as well as due to the unknown identity of some variables, it is impossible to add standards for all. Therefore, mostly, one internal standard (a known amount of a compound that is different from the metabolites present in the sample) is added to the sample mix, and the concentrations of all the metabolites present in the sample are then determined relative to the concentration of the internal standard. Internal standards are also used in targeted studies where there are a large number of targeted metabolites, or when a whole class of metabolites is targeted.

Since each of these approaches has its own advantages, disadvantages and unique applications, it is often necessary and recommended to use both targeted and untargeted methods of investigation, as was done in my study.

Metabolomics studies encompass various high-resolution identification techniques that survey for global changes in numerous metabolic pathways. These techniques include mass spectrometry (MS) and NMR spectroscopy, and are utilized in order to reliably measure a large fraction of the metabolites in a biological matrix. In this way, a more comprehensive view and in-depth knowledge of the complexity of biochemical processes can be obtained, and relationships between pathways and processes may be revealed (Castle *et al.*, 2006; Hendriks *et al.*, 2011). The identification steps using MS are often preceded by separation methods such as GC, LC or capillary electrophoresis (CE) (Steinfath *et al.*, 2008). These separation techniques reduce the mass spectral complexity and provide additional information on the physiochemical properties of the metabolites, which helps with their identification (Castle *et al.*, 2006; Little, 1999; Want *et al.*, 2010). Combining separation and identification techniques creates various hyphenated separation–identification platforms or approaches (including GC–MS, LC–MS, and CE–MS), each with its own strengths and weaknesses. GC–MS, for example, is a combined system where volatile and thermally stable compounds are first separated by gas chromatography, and then the eluting compounds are detected by electron impact mass spectroscopy. GC–MS allows for the rapid identification and quantification of volatile, non-polar metabolites (such as organic acids and metabolic derivatives of ethanol) with a high degree of sensitivity and chromatographic resolution. However, this platform is not without its analytical limitations and sources of variation, as will be discussed in more detail in Chapter 5.

Global analyses using MS and NMR can aid in the identification of changes in metabolic profiles associated with alcohol exposure and liver damage (Manna *et al.*, 2011). Both these methods were chosen to be used during my investigation to attempt to answer the research question stated in Chapter 1 — NMR as an untargeted approach, followed by GC–MS as a targeted method. Since the GC–MS investigation of my study covered a broad class of metabolites, namely the organic acids, it is the general opinion that “a semi-targeted approach” would be a more appropriate description of this method of analysis (Nielsen & Oliver, 2005). Using this approach, it is anticipated that the second method (GC–MS) would present further detail on

observations from the first method (NMR). This anticipated outcome is, however, not always the case, as will clearly be illustrated by my investigation on the effects of acute alcohol consumption (presented as two papers in Chapter 4 (section 4.5.2) and Chapter 5 (section 5.7), respectively, and discussed in Chapter 6).

In many metabolomics studies, the number of samples collected is often much smaller than the number of metabolites or variables measured, and simple visual inspection of the large amounts of metabolic data generated is generally not sufficient to complete the analysis. Thus, data are processed and interpreted using bioinformatics (univariate and multivariate statistical analyses) in order to extract the biologically relevant information (Dunn *et al.*, 2008; Goodacre *et al.*, 2007). Data processing and statistical analysis are key components of metabolomics studies, and thus these types of investigations require close collaboration between biochemists and statisticians. Therefore, parallel to my Ph.D. study, a Ph.D. study in Statistics is also being conducted on this research by a fellow student, Mari van Reenen. The aim of the Statistics study is based on the development of new bioinformatics tools for the analysis of metabolomics data, but analysis of my data on acute alcohol consumption also contributed to that Ph.D. study. Bioinformatics in metabolomics studies has become a vast field of expertise that is continuously expanding — providing opportunities for novel developments to uncover information from the complex metabolomics data sets generated. For the purpose of my study, I will focus only on the main lines of bioinformatics as required for its aims.

The data obtained from data generation and sample analysis are subjected to pre-processing steps (including noise reduction, baseline correction, normalization and scaling) to correct for disturbances due to environmental, or other, perturbations (Fukusaki & Kobayashi, 2005). Subsequently, statistical models are used to uncover information encapsulated in this pre-treated data. These methods of statistical analysis can be classified as either univariate or multivariate (Goodacre *et al.*, 2007):

- Univariate methods test individually for metabolites that are increased or decreased significantly between different groups. These include parametric methods, such as ANOVA (analysis of variance) and t-tests, for normally distributed data, and non-parametric methods when a normal distribution of the data cannot be assumed.
- Multivariate methods of analysis are often used in metabolomics due to the complicated linear and nonlinear relationships between metabolites in metabolomics data that must be elucidated in order to aid in the interpretation of the underlying properties of the data (Fukusaki & Kobayashi, 2005). Multivariate methods are further classified as supervised or unsupervised (Goodacre *et al.*, 2004; Goodacre *et al.*, 2007; Hendriks *et al.*, 2011):

- Supervised pattern recognition methods are powerful means that seek to transform the multivariate data from metabolite profiles into information with biological interest with the use of prior knowledge, such as the outcome of clinical tests — group membership information, as well as the metabolomic profile data, are used to build statistical models that attempt to explain the group separation. These models, including partial least squares discriminant analysis (PLS–DA) and ANOVA-simultaneous component analysis (ASCA), are also used as variable selection methods to reduce the number of variables for further interpretation, while still accounting for a sufficient percentage of the variation of interest, in this instance represented by group membership. This should provide an interpretable description of what, in biological terms, is the basis for the differentiation between groups.
- With unsupervised methods, such as principal component analysis (PCA) and hierarchical cluster analysis, no group membership is used in the statistical modelling of the metabolomic profile data. However, the group membership variable is used as a labelling variable to identify the natural grouping (perturbation patterns) that exists in the data, and that are described and/or revealed by these statistical models.
 - PCA is used to reduce the dimensionality of the data (while minimizing information loss), as well as to identify significant signals in noisy data. The application of PCA to a metabolomics data set provides two quantities from which useful information can be identified and indicated, namely the score and the loading: the score plot (which is limited to the most significant principal components) gives a visual image of the differences between samples from an all-around viewpoint; and the loading allows for the evaluation of the contribution that each metabolite makes to the total modelled information. The loading is useful to understand the differences in the levels of each metabolite between samples (Fukusaki & Kobayashi, 2005).
PCA results are also used to construct biplots — planar graphs that allow for visualization of information on both samples and variables of a data matrix simultaneously. For this purpose, the scores and loadings of two principal components (often the first two) are plotted as points and arrows, and the corresponding interpretation of the biplot enables the capturing of the main processes in the multivariate structure of the data set. The shorter the arrows, the more the ratio between the corresponding compounds tend to be constant throughout the data set, implying little to no ability to explain the observed variation. Longer arrows mean greater importance of the corresponding compound in explaining the patterns in the data set after projection to a lower dimensional subspace (Janečková *et al.*, 2012). In this representation, variable importance — the

proportion of variation explained by a given variable — is then relative to the amount of variation explained by the remaining variables.

- Clustering methods, such as hierarchical cluster analysis, are used to assess how similar a set of samples (or variables) are to one another, as well as to the rest of the data, based on their metabolite profiles. The graphical representation of these procedures is a two-dimensional ‘tree-like’ diagram, known as a dendrogram. The hierarchical methods, as used in this study, first cluster the most similar items, and then progressively add less similar items until all the items have been included in a single large cluster. This is reflected by the heights of the links between the clusters (samples) in the dendrogram — the lower the height of the link between the two clusters, the more similar the objects are (Goodacre *et al.*, 2007; Janečková *et al.*, 2012).

Statistical analyses are applied to provide low-dimensional representations (2D or 3D plots) of the high-dimensional multivariate metabolomics data. In this way, the structure in the data can be uncovered by visual inspection of the plots (Hendriks *et al.*, 2011). Statistical analyses can also be used to facilitate biological interpretation of the data. Such methods include (Goodacre *et al.*, 2007; Hendriks *et al.*, 2011):

- Network-inference methodology, which uses the knowledge that metabolites function in a network and are related to each other through pathways and other types of influence mechanisms.
- Correlation analysis that determines correlations, or partial correlations, between metabolites, which result from direct and indirect relations, and can be used to determine the correlation coefficients, as well as to construct correlation matrix maps.
- ASCA, which uses existing biological and experimental knowledge and prior information (such as information on treatment, time and whether a subject is a case or a control) to estimate the model parameters and direct the data analysis. This makes it possible to disentangle the effects of different factors (such as time and treatment) from a highly multivariate metabolomics data set.

More comprehensively, metabolomics can therefore be defined as an approach aimed at the non-biased identification and quantification of all small metabolites in a biological sample, using highly selective and sensitive analytical techniques, and analysed using bioinformatics (Dunn & Ellis, 2005), in order to investigate a biological phenomenon.

With regard to imbibing alcohol, we can broadly distinguish between chronic and acute consumption. The remainder of this chapter highlights the information uncovered through metabolomics investigations on these two modes of alcohol consumption.

3.2 Metabolomics and chronic alcohol consumption

Alcohol misuse and alcohol-related medical disorders are major, but neglected, public health problems worldwide and account for substantially elevated social and economic costs (Lieber, 1995). Because of this, the serious nature of the disease, and the availability of clinical material from patients suffering from chronic alcoholism, the effects of chronic alcohol consumption and alcoholism have been extensively researched and well described. Several metabolomics studies examining the effect of chronic alcohol consumption on metabolism, as well as research on various chronic disease states related to chronic alcohol consumption, have recently been reported, illustrating the value of a metabolomics approach to this area of research. The results from these, and other, metabolomics studies could serve to confirm findings from previous non-metabolomics studies, or they could suggest additional effects of chronic alcohol consumption on metabolism. Fourteen examples of these studies were examined, and are summarized or discussed in the next sections. Various analytical platforms, including GC–MS (1), LC–MS (7) and NMR (6), were involved in the studies examined. As no LC–MS metabolomics was used in my investigation, the seven examples of LC–MS metabolomics studies related to chronic alcohol consumption are summarized in Table 3–1, and the discussion of these studies will be limited to only this information. However, since GC–MS and NMR spectroscopy were used in my investigation, specific focus was placed on the single GC–MS study as well as the six investigations that used NMR.

3.2.1 Hyphenated MS studies

Zivkovic *et al.* used thin-layer and gas chromatography to perform lipid metabolomics analyses on liver samples from micropigs given a folate-deficient alcohol-containing diet (with or without S-adenosyl methionine (SAME) supplementation), as well as from control pigs. The purpose of this study was to estimate the flux of fatty acids through various pathways in order to determine the effect of alcohol consumption on key liver lipid metabolic pathways and their contribution to hepatic fat accumulation. The micropig is a suitable model of alcoholic liver disease (ALD) and alcoholic steatohepatitis, due to its propensity to develop alcoholism and liver injury (including fibrosis, inflammation, and other histological and biochemical markers of ALD) spontaneously with voluntary ethanol consumption. It has also been shown that micropigs develop fibrosis and inflammation, as well as other histological and biochemical markers of ALD over 12 months of ethanol consumption. SAME is the activated form of methionine, the ultimate precursor for cysteine, one of the three amino acids of GSH, and therefore its supplementation should help to alleviate alcohol-induced GSH depletion and its consequences (Lieber, 1997). Results from the study by Zivkovic *et al.* showed that, compared to the controls, the alcoholic pigs had increased levels of hepatic triglycerides (which was at least partially attenuated with SAME supplementation), increased flux through the lipid synthesis pathway (catalysed by stearyl-CoA

desaturase), and reduced flux through the lipid export (catalysed by phosphatidylethanolamine-N-methyltransferase) and fatty acid elongation pathways. These results show that both increased lipid synthesis and altered phospholipid metabolism (e.g. lipid export) contribute to alcohol-induced hepatic fat accumulation (which often leads to alcoholic steatohepatitis) (Zivkovic *et al.*, 2009).

Table 3-1: Examples of recent metabolomics studies (performed using various LC–MS techniques) on the effect of chronic alcohol consumption on metabolism, as well as of chronic disease states related to chronic alcohol consumption.

Disease state	Study method and aim	Reference
Oxidative stress-induced alcoholic liver damage	High-performance liquid chromatography (HPLC) analysis was performed on liver extracts from mice challenged with binge-like ethanol administration, to investigate the association between changes in hepatic concentrations of sulphur-containing substances and the induction of oxidative stress-induced alcoholic liver damage, as well as the protective effect of dietary betaine against ethanol-induced liver injury.	Kim <i>et al.</i> , 2008
ALD	Mass spectrometry-based analyses (using a combination of ultra-performance liquid chromatography coupled with electrospray ionization quadrupole time-of-flight mass spectrometry (UPLC–ESI–QTOFMS) and chemometrics) were performed on alcohol-fed <i>Ppara</i> -null mice, to examine the metabolic changes associated with ALD, as well as to unravel the molecular mechanism of the disease pathogenesis and potentially to identify biomarkers of the disease.	Manna <i>et al.</i> , 2010
	Reversed-phase gradient UPLC–ESI–QTOFMS analyses were performed on urine samples from <i>Ppara</i> -null mice (as well as their wild-type counterparts) treated with chronic alcohol consumption, to study the effect of genetic background on the ALD-associated urinary metabolome, with the aim of possibly identifying a genetic background-independent noninvasive ALD biomarker.	Manna <i>et al.</i> , 2011
	Hepatic transcriptomic and metabolomic analyses (microarray analysis and ultra-high-performance liquid chromatography quadrupole time-of-flight mass spectrometry (UPLC–QTOFMS)) were performed on liver samples from <i>Ppara</i> -null and WT mice, to examine the changes in gene expression and metabolites associated with pathological changes during chronic alcohol consumption in order to better understand ALD development and the dynamic regulation of <i>Ppara</i> activity.	Li <i>et al.</i> , 2014
	High-throughput UPLC–ESI–QTOFMS analyses (combined with various pattern recognition techniques) were performed on human urine samples from confirmed ALD sufferers and controls, to identify potential urinary biomarkers for the early detection of ALD.	Liang <i>et al.</i> , 2015
Acute alcoholic hepatitis	Untargeted ultra-high-performance liquid chromatography coupled with electrospray ionization tandem mass spectrometry (UPLC–ESI–MSMS) analyses (combined with various statistical analyses) were performed on serum samples from patients with severe acute alcoholic hepatitis (AAH) and patients with alcoholic cirrhosis, to compare the metabolic profiles of these two groups in order to characterize the perturbations in the metabolic pathways associated with severe AAH in humans.	Rachakonda <i>et al.</i> , 2014
Hepatocellular carcinoma (HCC)	UPLC–QTOFMS analyses were performed on serum samples from nude mice, mice with alcohol-induced liver injury (achieved by daily administration of grain liquor), and mice bearing a HCC xenograft (achieved by subcutaneous inoculation of HepG2 cells), to investigate the disease-associated metabolite changes responsible for alcohol-induced liver injury and its progression to HCC. The acquired data were analysed by PCA and orthogonal partial least squares discriminant analysis (OPLS–DA) to identify potential disease-specific biomarkers.	Li <i>et al.</i> , 2011

3.2.2 NMR studies

3.2.2.1 Experimental animal studies

Animal, mostly mouse, models that mimic human alcoholic liver injury and ALD were used in some of the LC–MS studies summarized in Table 3–1. Nuclear peroxisome proliferator-activated receptor alpha (*Ppara*)-null mice are often used, since earlier studies have suggested that the activity of *Ppara* protects alcohol-fed mice against ALD — *Ppara*-null mice develop more pronounced hepatic changes than wild-type (WT) mice after chronic alcohol exposure. Also, despite considerable differences in their genomes, the mouse is an excellent model for elucidating the molecular events leading to the development of human diseases. The remarkable similarities in the histopathology of alcohol-treated *Ppara*-null mice to hallmarks of human ALD, and the fact that *Ppara* expression and activity in human liver are less than one tenth of those in WT mouse liver, makes *Ppara*-null mice an excellent model for studying system level (epigenomic, transcriptomic, proteomic, and metabolomic) changes associated with ALD (Li *et al.*, 2014; Manna *et al.*, 2010).

Animal models were also used in two of the NMR studies examined in the literature:

1. Bradford *et al.* performed NMR spectroscopy and electrospray infusion/Fourier transform ion cyclotron resonance–mass spectrometry (ESI/FTICR–MS) analyses on urine and liver extracts from mice on a high-fat, corn oil-based, alcohol-containing liquid diet, to assess the biochemical and molecular changes that occur during disease progression towards alcohol-induced liver injury. Mice on this diet developed liver injury that was consistent with the human disease, starting with steatosis, hepatic inflammation and necrosis, progressing to steatohepatitis and eventually fibrosis.
They observed: (1) hypoxia and subsequent activation of glycolysis, increased lipid peroxidation and ketogenesis, and reduced citric acid cycle and gluconeogenesis; (2) increased excretion of lactate, tyrosine and threonine; and (3) decreased excretion of taurine, indicating increased oxidative stress and subsequent depleted GSH.
Additionally, important similarities (including increased ethylglucuronide, n-acetylglutamine and n-acetylglycine excretion, and similar changes in liver fatty acid profiles) were observed when the results were compared with those from a human study performed on 186 male drinkers (Simon *et al.*, 1996), which supports the human relevance of this mouse model (Bradford *et al.*, 2008).
2. Fernando *et al.* performed NMR spectroscopy analyses (followed by multivariate statistical analyses) on lipids and phospholipids extracted from plasma and liver samples from alcohol-fed rats and control rats, in order to study the altered lipid metabolome (lipidome) and to

understand the mechanisms and lipid pattern of early-stage alcohol-induced fatty liver (hepatic steatosis).

They observed: (1) significant fatty infiltration in stained liver sections, but no inflammation or oxidative stress, suggesting that hepatic steatosis precedes these symptoms; (2) significant differences in the lipidomes, and concentrations of various phospholipids, in the samples from the alcohol-fed rats compared to the control rats; and (3) altered cholesterol, triglycerides, fatty acids, and phospholipids metabolism, which contributes to the development of fatty liver (Fernando *et al.*, 2010).

However, despite animal models being appropriate subjects for studies involving induced chronic alcoholism, these models are not suitable to be used in the assessment of the effects of acute alcohol consumption for various ethical and logistical reasons, including: (1) obtaining ethical approval for such a study — where participation is involuntary and experimental subjects cannot provide informed consent — would be difficult or impossible; (2) effective administration of a single large acute dose of alcohol; and (3) frequent/periodic sample collection. This necessitates the use of human experimental subjects to perform studies on acute alcohol consumption.

3.2.2.2 Global metabolite profile of chronic alcoholism

NMR spectroscopy has also been used in several studies involving human patients and samples to help elucidate the global metabolite profiles associated with chronic alcohol consumption and the related disease states. Figure 3–2 gives an example of such results, clearly showing how statistical analyses (A) of NMR spectroscopy data can be used to differentiate between different alcohol-related disease states (B and C).

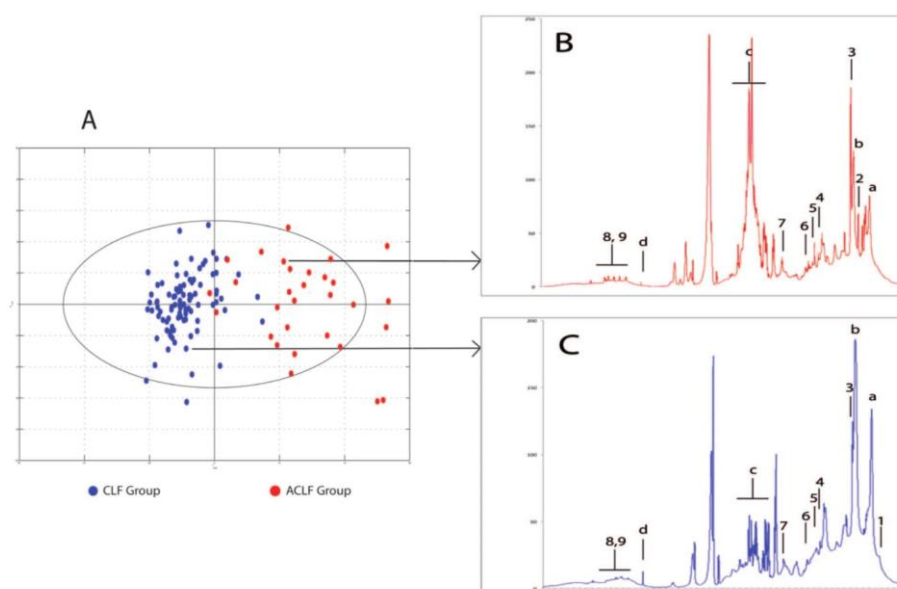


Figure 3-2: OPLS score plot and spectra of chronic liver failure (CLF) and acute-on-chronic liver failure (ACLF) patients [reproduced from Amathieu *et al.*, 2014]. A: Score plot separating CLF (blue) and ACLF (red) patients. B and C: Representative spectra showing the metabolic differences between patients with CLF or ACLF. Legends: a, -CH₃ of fatty acid; b, -CH₂ of fatty acid; c, glucose (C) and hydroxy-ethyl starch (B); d, fumaric acid; 1, cholesterol; 2, hydroxybutyrate; 3, lactate; 4, acetoacetate; 5, pyruvate; 6, glutamine; 7, creatinine; 8 & 9, aromatic amino acids (phenylalanine and tyrosine).

Three of these human NMR metabolomics studies were performed by the same group of researchers, and one by another group:

1. In 2011, Amathieu *et al.* performed ¹H-NMR spectroscopy analyses (followed by statistical analyses) on serum samples of patients with alcoholic cirrhosis, in order to assess the severity of chronic liver failure (CLF) based on their metabolomic profiles. Compared to patients with mild CLF, patients with severe CLF had: (1) elevated levels of metabolites such as glucose (indicative of increased severity of liver disease), amino acids (such as leucine, isoleucine, glutamine, and methionine), lactate and pyruvate (possibly due to impaired pyruvate-dehydrogenase activity), butyrate, citrate (indicative of impaired Krebs cycle activity), and creatinine; and (2) decreased levels of high-density lipoprotein, phosphocholine, alanine, and ketone bodies (mainly β-hydroxybutyrate and acetate) (Amathieu *et al.*, 2011).
2. A follow-up NMR metabolomics study was performed by Amathieu *et al.* in 2014 on serum samples from patients with either alcoholic cirrhosis or cirrhosis as well as acute-on-chronic

liver failure, in order to compare the metabolomic profiles of these two groups, and to identify the metabolic changes related to acute-on-chronic liver failure (ACLF).

Various metabolic changes related to ACLF were identified, including: (1) increased levels of lactate (a biomarker for tissue hypoxia and necrosis), pyruvate, ketone bodies (indicating increased use of lipids as source of energy due to alcohol-induced depressed glycolysis), glutamine and glutamate (due to impaired urea metabolism and consequent decreased ammonia detoxification), phenylalanine and tyrosine (early markers of liver inflammation and injury), and creatinine (indicating advanced liver failure); and (2) decreased high-density lipoproteins, which indicates poor hepatic function. The changes in these metabolites could possibly be used to differentiate between patients with cirrhosis with or without ACLF (Amathieu *et al.*, 2014).

3. Nahon *et al.* performed ^1H -NMR spectroscopy analyses (combined with various statistical techniques) on serum samples from cirrhotic patients with and without hepatocellular carcinoma (HCC), in order to assess their metabolomic profiles and to identify a metabolic fingerprint specific to HCC in patients with alcoholic cirrhosis.

Compared to the cirrhosis-only patients, patients with HCC had: (1) significantly increased concentrations of several metabolites, including glutamate, acetate and N-acetyl glycoproteins; and (2) decreased levels of high-density lipoproteins, unsaturated lipids, and glutamine (possibly due to increased glutamine consumption by tumour cells for energy production) (Nahon *et al.*, 2012).

4. Qi *et al.* performed ^1H -NMR spectroscopy analyses on serum samples of hepatitis B virus (HBV)-infected cirrhosis patients, alcoholic cirrhosis patients and controls, in order to compare their metabolic profiles. The acquired data were analysed by applying various statistical analyses potentially to identify specific biomarkers associated with the diseases.

Compared to controls, both alcoholic cirrhosis and HBV-infected cirrhosis patients showed: (1) increased levels of α -glucose, β -glucose, tyrosine, histidine, phenylalanine, pyruvate, succinate, trimethylamine, citrate, and 3-hydroxybutyrate; and (2) decreased levels of low-density lipoproteins, very low-density lipoproteins, lipids, valine, N-acetyl glycoprotein signals, lysine, leucine, glycerophosphocholine/phosphocholine, isoleucine, lactate and acetate.

Compared with alcoholic cirrhosis patients, HBV-infected cirrhosis patients showed: (1) higher levels of creatine (due to necrotic hepatocytes) and isobutyrate (indicating diminished liver function); and (2) lower levels of acetoacetate (due to impaired hepatic energy metabolism), as well as lower levels of glutamate and glutamine (two possible sensitive biomarkers for alcoholic cirrhosis) (Qi *et al.*, 2012).

3.2.2.3 New insights from chronic alcoholism studies

All of the studies discussed above refer to the new insights gained through the use of metabolomics and NMR spectroscopy, or to the benefit and value of this approach to future research:

1. The metabolomic NMR analyses by Amathieu *et al.* provided new insights into the metabolic processes related to the severity of hepatic function impairment due to cirrhosis. Their study also emphasized the usefulness of a metabolomics approach to diagnose and assess the severity of CLF (Amathieu *et al.*, 2011).
2. The application of a metabolomics approach in the follow-up by Amathieu *et al.*, provided complete and multivariate information on the metabolomic changes related to ACLF, improved the classification of patients with or without ACLF, and contributed to an increased understanding of the involvement of the biological pathways in ACLF (Amathieu *et al.*, 2014).
3. The study by Nahon *et al.* provided new insights into the metabolic processes related to liver cancer, which can be useful for determining prognosis and treatment methods in these patients (Nahon *et al.*, 2012).
4. In their study, Qi *et al.* suggested that ^1H -NMR and metabolic profiling of serum samples may provide a new way to explore the pathogenesis of HBV-infected and alcoholic cirrhosis patients, which could aid in their diagnosis (Qi *et al.*, 2012).

Conclusion: The observations, results and conclusions from these metabolomics studies on chronic alcohol consumption and the related pathological consequences highlight the value of metabolomics and NMR spectroscopy and their ability to provide new insights into chronic alcohol consumption and the related diseases. However, a limitation of all of these studies is that they give only “snapshot” information (that is, a static image of the metabolic profile of the disease), at a given point in time, during the development of the disease. The information from these studies is valuable, but would have increased impact if they could be designed as longitudinal (time-dependent) studies which would provide a dynamic component to complement the results from the investigations discussed above. Therefore, my investigation followed a similar approach to the investigations discussed above, but with the additional inclusion of a time factor (Chapter 4). It is fair to assume that this addition would add value to alcohol-related studies, and that the scientific results would be equally, or even more, informative. Designing a study implementing a longitudinal approach is, however, not easy. Since acute alcohol consumption can be more easily controlled and is more homogeneous than chronic alcohol consumption, designing a longitudinal study to investigate the effects of acute alcohol consumption seems more feasible to achieve. Therefore, it was decided to investigate acute, rather than chronic, alcohol consumption in the present investigation.

3.3 Metabolomics and acute alcohol consumption

High-risk drinking patterns, such as binge drinking, result in an increased risk of chronic alcohol abuse and the development of alcohol dependence — binge drinking at least once a week is an early quantifiable marker for alcohol use disorders (Zakhari & Li, 2007). The increasing prevalence of worldwide alcohol consumption highlights the importance of research in this field — knowing the signs and identifying biomarkers of acute alcohol consumption should help with identification of the problem before it becomes too serious. The clinical consequences and metabolic perturbations associated with acute alcohol consumption have been addressed by numerous researchers over the years. Most of this research was, however, based on traditional methodologies, which focused on only selected perturbations and metabolic pathways (Chapter 2).

Metabolomics approaches have been used to examine various physiological processes and diseases (Li *et al.*, 2011), and would be equally applicable and valuable to studies on the effects of acute alcohol consumption on metabolism for several reasons:

- Metabolomics makes for an appropriate solution to the problems associated with studies based on traditional methodologies, since it is a holistic approach, taking the entire metabolic profile into consideration. Such an approach allows for a more comprehensive analysis of the metabolome (the global quantitative collection of all metabolites in an organism at a given time) (Dunn *et al.*, 2008), and would permit more in-depth analysis and increased understanding of the metabolic perturbations associated with acute alcohol consumption.
- Metabolites are the end-products of a cascade of interactions involving numerous biological molecules (including DNA, RNA, and proteins). Thus, the metabolome is seen as the most downstream result of gene expression, and represents the closest molecular level description of the physiological states and activities of cells. In principle, any physiological perturbation is expected to be associated with characteristic changes in the metabolome, and these changes are expected to be amplified relative to changes in the transcriptome and the proteome. Therefore, metabolomics research (in general and also related to acute alcohol consumption) could lead to new insights and discoveries that could have been missed using more exclusive methods of investigation such as transcriptomics and proteomics. (Dettmer *et al.*, 2007; Dunn *et al.*, 2008; Goodacre *et al.*, 2004; Manna *et al.*, 2010; Quéro *et al.*, 2014).
- The application of metabolomics approaches could add value to alcohol-associated research, since it allows for the identification and analysis of an increased number of

metabolites within different metabolite classes. This permits researchers to gain a greater understanding of the distinct biochemical processes with which these metabolites are associated.

- Metabolomics analyses of the metabolites produced from the various alcohol-induced perturbed metabolic pathways may also help to determine the contributions of different metabolites and metabolic pathways to the development of alcohol-related tissue damage and disease progression. This could aid in the identification of biomarkers of alcohol abuse, as well as in the development of effective interventions, and perhaps even individualized treatment regimes (Harrigan *et al.*, 2008).
- The altered metabolism of metabolites from one class can influence the metabolism of metabolites from another class. Indeed, the concentrations of a small but significant number of metabolites are known to be highly interrelated when measurements of biological replicates are conducted (Steuer, 2006). Metabolomics maximizes the number of metabolites that can be measured in a single experiment, thereby aiding in the elucidation of these types of interactions (Harrigan *et al.*, 2008). The relationships between the concentrations of metabolites can then be expressed using covariance or correlation analyses.
- The highly sensitive and specific analysis platforms used in metabolomics studies make it possible to identify and measure the often very subtle concentration changes of certain metabolites that occur as a result of acute alcohol consumption.
- In addition, with the aid of bioinformatics particular metabolites of interest can be isolated, identified and credited as the variables responsible for the greatest variation between control and experimental groups.

These advantages and applications of metabolomics to acute alcohol research serve as justification for undertaking the study presented here. However, the administration of alcohol to experimental participants presents the researcher with several challenges, of which two aspects require special emphasis:

1. Development of a well-designed (yet executable) experimental design (discussed comprehensively in Chapter 4).
2. Ethics.

The conduct during any research process should be ethically sound in order to protect North-West University (NWU), the researcher, the research team, research participants and the environment. At the NWU, researchers must apply for ethical clearance from the

relevant research ethics committee (REC), and this application must be approved before any research may be conducted. Proof of ethical approval comes in the form of an NWU ethics certificate signed by the chairperson of the **Institutional Research Ethics Committee (IRERC)** (NWU, 2017). The NWU provides a “Framework for the Management of Research Ethics at NWU”, which gives an indication of how research should be managed and conducted so that it is in line with the ethical principles adopted by the university. This framework, which is available online at www.nwu.ac.za/research-support/ethics, includes a “Policy Statement”, a “Code of Conduct for Researcher at the NWU”, and “Rules for the Management of Research Ethics at the NWU”. This policy complies with the international convention on the ethical conduct of research, with special reference to human participants as cases and controls. The code of conduct must be signed by the researcher upon approval of a research ethics application before the start of the study, and strictly adhered to during the investigation. Only upon receiving a research ethics certificate, as well as signing the code of conduct, may the researcher continue with his/her specific research project.

Since my study involved exposing individuals to an intervention, in this case alcohol, with known potentially serious mental, societal and health risks, several ethical considerations needed to be addressed during the design of this study. These include: (1) the number of participants required for the study, since individuals’ attitudes to their consumption habits is a private issue; (2) the number of urine samples required to be collected over the period of the intervention, since sufficient numbers are needed for satisfactory analysis and interpretation, but taking into account that voluntary urine specimens can only be provided every so often; (3) the total duration of the intervention period, since all participants were in a fasting state and could not remain so indefinitely; and (4) the volume and type of alcohol to be consumed, as well as the maximum period and method of consumption. For these reasons, this study required the application for ethical approval from a relevant REC, in this case the Health Research Ethics Committee (HREC), which is registered with the National Health Research Ethics Council (NHREC) and handles all research ethics applications relating to research on humans and health, from the Faculty of Health Sciences and the Faculty of Natural Sciences on the Potchefstroom Campus of NWU (NWU, 2017). More information on the HREC can be found on their webpage at <http://health-sciences.nwu.ac.za/healthethics/>.

The process of application of ethical approval for a research study involving human (or animal) subjects or samples is extensive and involves many steps. A standard operating procedure (SOP) for the application process of research ethics approval is available online at <http://health-sciences.nwu.ac.za/healthethics/HREC/application-notification-forms>. The purpose of the SOP is to provide researchers with a clear, systematic procedure to follow when applying for ethical approval, and covers the full application process to obtain

research ethics approval before research is conducted, permission for amendments and the monitoring process during research. As of June 2016, all researchers applying for ethical approval with the HREC are also required to provide proof of formal training in research ethics within the past three years. This proof can be in the form of a letter of attendance at a two-day training course of the Ethics Office of the Faculty of Health Sciences, entitled "The Basics of Health Research Ethics", or a certificate from an appropriate Training and Resources in Research Ethics Evaluation (TRREE) online training course.

Ethical approval for the experimental protocol of my study was, however, already obtained in 2012, before HREC was established in 2014, by the NWU Ethics Committee (ethical approval number: NWU-00045-12-S1). All participants also provided informed written consent to the research protocol. Copies of the ethical approval certificate and informed consent form are included in the annexure at the end of this thesis.

REFERENCES

- Amathieu, R., Nahon, P., Triba, M., Bouchemal, N., Trinchet, J-C., Beaugrand, M., Dhonneur, G. & Le Moyec, L. 2011. Metabolomic approach by ^1H NMR spectroscopy of serum for the assessment of chronic liver failure in patients with cirrhosis. *Journal of proteome research*, 10(7):3239–3245.
- Amathieu, R., Triba, M.N., Nahon, P., Bouchemal, N., Kamoun, W., Haouache, H., Trinchet, J-C., Savarin, P., Le Moyec, L. & Dhonneur, G. 2014. Serum ^1H -NMR metabolomic fingerprints of acute-on-chronic liver failure in intensive care unit patients with alcoholic cirrhosis. *PLOS ONE*, 9(2):e89230.
- Bernini, P., Bertini, I., Luchinat, C., Nincheri, P., Staderini, S. & Turano, P. 2011. Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. *Journal of biomolecular NMR*, 49(3–4):231–243.
- Bradford, B.U., O'Connell, T.M., Han, J., Kosyk, O., Shymonyak, S., Ross, P.K., Winnike, J., Kono, H. & Rosyn, I. 2008. Metabolomic profiling of a modified alcohol liquid diet model for liver injury in the mouse uncovers new markers of disease. *Toxicology and applied pharmacology*, 232(2):236–243.
- Bundy, J.G., Davey, M.P. & Viant, M.R. 2009. Environmental metabolomics: a critical review and future perspectives. *Metabolomics*, 5(1):3–21.
- Castle, A.L., Fiehn, O., Kaddurah-Daouk, R. & Lindon, J.C. 2006. Metabolomics standards workshop and the development of international standards for reporting metabolomics experimental results. *Briefings in bioinformatics*, 7(2):159–165.
- Dettmer, K., Aronov, P.A. & Hammock, B.D. 2007. Mass spectrometry-based metabolomics. *Mass spectrometry reviews*, 26(1):51–78.
- Dunn, W.B. & Ellis, D.I. 2005. Metabolomics: current analytical platforms and methodologies. *Trends in analytical chemistry*, 24(4):285–294.
- Dunn, W.B., Broadhurst, D., Ellis, D.I., Brown, M., Halsall, A., O'Hagan, S., Spasic, I., Tseng, A. & Kell, D.B. 2008. A GC–TOF–MS study of the stability of serum and urine metabolomes during the UK biobank sample collection and preparation protocols. *International journal of epidemiology*, 37(suppl.1):i23–i30.

- Fernando, H., Kondraganti, S., Bhopale, K.K., Volk, D.E., Neerathilingam, M., Kaphalia, B.S., Luxon, B.A., Boor, P.J. & Ansari, G.A.S. 2010. ^1H and ^{31}P NMR lipidome of ethanol-induced fatty liver. *Alcoholism: clinical and experimental research*, 34(11):1937–1947.
- Fiehn, O. 2001. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and functional genomics*, 2(3):155–168.
- Fiehn, O., Robertson, D., Griffin, J., Van der Werf, M., Nikolau, B., Morrison, N., Sumner, L.W., Goodacre, R., Hardy, N.W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., Van Ommen, B., Lindon, J.C. & Sansone, S-A. 2007. The metabolomics standards initiative (MSI). *Metabolomics*, 3(3):175–178.
- Fukusaki, E. & Kobayashi, A. 2005. Plant metabolomics: potential for practical operation. *Journal of bioscience and bioengineering*, 100(4):347–354.
- Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G. & Kell, D.B. 2004. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends in biotechnology*, 22(5):245–252.
- Goodacre, R., Broadhurst, D., Smilde, A.K., Kristal, B.S., Baker, J.D., Beger, R., Bessant, C., Connor, S., Capuani, G., Craig, A. & Ebbels, T. 2007. Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics*, 3(3):231–241.
- Harrigan, G.G., Maguire, G. & Boros, L. 2008. Metabolomics in alcohol research and drug development. *Alcohol research & health*, 31(1):26–35.
- Hendriks, M.M.W.B., Van Eeuwijk, F.A., Jellema, R.H., Westerhuis, J.A., Reijmers, T.H., Hoefsloot, H.C.J. & Smilde, A.K. 2011. Data-processing strategies for metabolomics studies. *Trends in analytical chemistry*, 30(10):1685–1698.
- Janečková, H., Hron, K., Wojtowicz, P., Hlídková, E., Barešová, A., Friedecký, D., Žídková, L., Horník, P., Behúlová, D., Procházková, D., Vinohradská, H., Pešková, K., Bruheim, P., Smolka, V., Št'astná, S. & Adam, T. 2012. Targeted metabolomic analysis of plasma samples for the diagnosis of inherited metabolic disorders. *Journal of chromatography A*, 1226:11–17.
- Kim, S.J., Jung, Y.S., Kwon, D.Y. & Kim, Y.C. 2008. Alleviation of acute ethanol-induced liver injury and impaired metabolomics of S-containing substances by betaine supplementation. *Biochemical and biophysical research communications*, 368(4):893–898.

- Koek, M.M., Jellema, R.H., Van der Greef, J., Tas, A.C. & Hankemeier, T. 2011. Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. *Metabolomics*, 7(3):307–328.
- Kristensen, M., Engelsen, S.B. & Dragsted, L.O. 2012. LC–MS metabolomics top-down approach reveals new exposure and effect biomarkers of apple and apple-pectin intake. *Metabolomics*, 8:64–73.
- Li, H-H., Tyburski, J.B., Wang, Y-W., Strawn, S., Moon, B-H., Kallakury, B.V.S., Gonzalez, F.J. & Fornace Jr., A.J. 2014. Modulation of fatty acid and bile acid metabolism by peroxisome proliferator-activated receptor α protects against alcoholic liver disease. *Alcoholism: clinical and experimental research*, 38(6):1520–1531.
- Li, S., Liu, H., Jin, Y., Lind, S., Caic, Z. & Jiang, Y. 2011. Metabolomics study of alcohol-induced liver injury and hepatocellular carcinoma xenografts in mice. *Journal of chromatography B*, 879:2369–2375.
- Liang, Q., Wang, C., Li, B. & Zhang, A-H. 2015. Metabolomics of alcoholic liver disease: a clinical discovery study. *RSC advances*, 5(98):80381–80387.
- Lieber, C.S. 1995. Medical disorders of alcoholism. *New England journal of medicine*, 333(16):1058–1065.
- Lieber, C.S. 1997. Ethanol metabolism, cirrhosis and alcoholism. *Clinica chimica acta*, 257(1):59–84.
- Little, J.L. 1999. Artifacts in trimethylsilyl derivatization reactions and ways to avoid them. *Journal of chromatography A*, 844(1):1–22.
- Manna, S.K., Patterson, A.D., Yang, Q., Krausz, K.W., Li, H., Ilde, J.R., Fornace Jr., A.J. & Gonzalez, F.J. 2010. Identification of noninvasive biomarkers for alcohol-induced liver disease using urinary metabolomics and the *Ppara*-null mouse. *Journal of proteome research*, 9(8):4176–4188.
- Manna, S.K., Patterson, A.D., Yang, Q., Krausz, K.W., Ilde, J.R., Fornace Jr., A.J. & Gonzalez, F.J. 2011. UPLC–MS-based urine metabolomics reveals indole-3-lactic acid and phenyllactic acid as conserved biomarkers for alcohol-induced liver disease in the *Ppara*-null mouse model. *Journal of proteome research*, 10(9):4120–4133.

Nahon, P., Amathieu, R., Triba, M.N., Bouchemal, N., Nault, J-C., Ziol, M., Seror, O., Dhonneur, G., Trinchet, J-C., Beaugrand, M. & Le Moyec, L. 2012. Identification of serum proton NMR metabolomic fingerprints associated with hepatocellular carcinoma in patients with alcoholic cirrhosis. *Clinical cancer research*, 18(24):6714–6722.

Nielsen, J. & Oliver, S. 2005. The next wave in metabolome analysis. *Trends in biotechnology*, 23(11):544–546.

NWU (North-West University). 2017. Research support — ethics. <http://www.nwu.ac.za/research-support/ethics> Date of access: 10 Oct. 2017.

Qi, S., Tu, Z., Ouyang, X., Wang, L., Peng, W., Cai, A. & Dai, Y. 2012. Comparison of the metabolic profiling of hepatitis B virus-infected cirrhosis and alcoholic cirrhosis patients by using ¹H NMR-based metabonomics. *Hepatology research*, 42(7):677–685.

Quéro, A., Jousse, C., Lequart-Pillon, M., Gontier, E., Guillot, X., Courtois, B., Courtois, J. & Pau-Roblot, C. 2014. Improved stability of TMS derivatives for the robust quantification of plant polar metabolites by gas chromatography–mass spectrometry. *Journal of chromatography B*, 970:36–43.

Rachakonda, V., Gabbert, C., Raina, A., Bell, L.N., Cooper, S., Malik, S. & Behari, J. 2014. Serum metabolomic profiling in acute alcoholic hepatitis identifies multiple dysregulated pathways. *PLOS ONE*, 9(12):e113860. doi:10.1371/journal.pone.0113860.

Robertson, D.G., Watkins, P.B. & Reily, M.D. 2011. Metabolomics in toxicology: preclinical and clinical applications. *Toxicological sciences*, 120(S1):S146–S170.

Simon, J.A., Fong, J., Bernert Jr., J.T. & Browner, W.S. 1996. Relation of smoking and alcohol consumption to serum fatty acids. *American journal of epidemiology*, 144(4):325–334.

Spratlin, J.L., Serkova, N.J. & Eckhardt, S.G. 2009. Clinical applications of metabolomics in oncology: a review. *Clinical cancer research*, 15(2):431–440.

Steinfath, M., Groth, D., Lisec, J. & Selbig, J. 2008. Metabolite profile analysis: from raw data to regression and classification. *Physiologia plantarum*, 132(2):150–161.

Steuer, R. 2006. On the analysis and interpretation of correlations in metabolomic data. *Briefings in bioinformatics*, 7(2):151–158.

Vinayavekhin, N. & Saghatelian, A. 2010. Untargeted metabolomics. *Current protocols in molecular biology*, 90(30.1):1–24.

Want, E.J., Wilson, I.D., Gika, H., Theodoridis, G., Plumb, R.S., Shockcor, J., Holmes, E. & Nicholson, J.K. 2010. Global metabolic profiling procedures for urine using UPLC–MS. *Nature protocols*, 5(6):1005–1018.

Weljie, A.M., Newton, J., Mercier, P., Carlson, E. & Slupsky, C.M. 2006. Targeted profiling: quantitative analysis of ^1H NMR metabolomics data. *Analytical chemistry*, 78(13):4430–4442.

Zakhari, S. & Li, T-K. 2007. Determinants of alcohol use and abuse: impact of quantity and frequency patterns on liver disease. *Hepatology*, 46(6):2032–2039.

Zhang, A., Sun, H. & Wang, X. 2012. Serum metabolomics as a novel diagnostic approach for disease: a systematic review. *Analytical and bioanalytical chemistry*, 404(4):1239–1245.

Zhou, B., Xiao, J.F., Tuli, L. & Ressom, H.W. 2012. LC–MS-based metabolomics. *Molecular biosystems*, 8(2):470–481.

Zivkovic, A.M., German, J.B., Esfandiari, F. & Halsted, C.H. 2009. Quantitative lipid metabolic changes in alcoholic micropigs with fatty liver disease. *Alcoholism: clinical and experimental research*, 33(4):751–758.

CHAPTER 4 NMR METABOLOMICS OF ACUTE ALCOHOL CONSUMPTION

4.1 Background on intervention studies

A primary objective of metabolomics is to study the metabolism of an organism. Traditional metabolomics studies, such as fingerprinting or metabolic profiling, provide metabolite profiles of living systems at a single point in time. This approach generates important and valuable information, but does not capture the dynamics of a system. The key to investigate and gain insight into dynamic phenomena is through the analysis of a time series, often through the application of intervention studies. This approach provides the background for the experimental design used here.

The application of metabolomics to intervention studies greatly enhances the holistic understanding of the effects of the interventions (e.g. a dietary modification or a consumed substance) on metabolic pathways (Gibney *et al.*, 2005; Wishart, 2008a). The analysis of time series data is a fruitful method for the development of novel bioinformatics models and approaches for the analysis of metabolomics data (Jansen *et al.*, 2004; Rubingh *et al.*, 2009; Smilde *et al.*, 2005; Verouden *et al.*, 2009). Additionally, the information uncovered through these types of analyses provides the biochemist the opportunity for a better understanding of: (1) the commonalities in the perturbations observed in response to the interventions; (2) indications of inter-individual variability; and (3) the return to homeostasis (in the case of a single acute intervention, as studied here).

Metabolomics data, generated at a fixed time point, are represented as a multivariate matrix of controls and cases (rows) measured over the metabolites that reflect the perturbation under investigation (columns). Such data are processed through multivariate statistical methods — using mostly unsupervised PCA, and supervised methods such as PLS–DA, orthogonal projection to latent-structure (OPLS), and their variants — to identify metabolites that differ significantly among the test groups studied (Lindon *et al.*, 2001).

Time-dependent intervention studies (such as the study reported here), however, require more advanced statistical methods for data analysis. Data sets from time-dependent intervention studies are complex, as these investigations aspire to measure multiple metabolites in a biofluid, obtained from several experimental subjects, collected at different points in time, and subjected to interventions from different consumed substances. These studies call for methods of data analysis specifically designed for longitudinal (time-dependent), multi-subject (data from

several experimental participants), multi-group (intervention studies) and multivariate data (Antti *et al.*, 2002; Hendriks *et al.*, 2011; Smilde *et al.*, 2005).

Also, it is well-known that large, individual, within-subject variation in urinary organic acid composition occurs in individuals from the general healthy population, like the experimental subjects who participated in the present study. This variation can be due to several factors, some demographic (such as gender, age and ethnicity) and others environmental (such as diet and lifestyle) (Chalmers *et al.*, 1976; Pasikanti *et al.*, 2008; Zelena *et al.*, 2009). In human intervention studies, this individual variation tends to dominate the often less pronounced metabolic perturbations due to the intervention. A reasonably large number of cases/subjects is therefore mostly required, as well as a sufficient number of time points to cover the effect of the intervention.

The cross-over experimental design used for my study takes the complex aspects mentioned above into account in order to theoretically generate reliable and useful data that can be interpreted as relevant biological information. As this design generated several hundred samples, a different approach from the traditional multivariate methods for processing the data is required, as well as more extensive validation of the results than for metabolomics studies.

4.2 Experimental design for an intervention study

Designing an intervention study on acute alcohol consumption requires consideration of several experimental aspects, including:

1. Foremost were the ethical considerations that needed to be addressed when exposing individuals to an intervention with known serious mental, societal and health risks – this has, however, been addressed in Chapter 3.
2. Measuring an acute effect of alcohol consumption requires the intake of a relatively high dose of alcohol (1.5 g per kg body mass) within a short period of time (not exceeding 5 minutes). For this purpose, commercial alcohol (vodka) was selected.
3. For many people, including those that participated in this study, the typical volume of vodka used (~100 mL) would be difficult to consume neat in the restricted time period. To overcome this, the vodka was imbibed together with a consumer-friendly vehicle (500 mL lemon-flavoured water). This, however, is an intervention in itself, which then needed to be accounted for.
4. Finally, it is well-known that there is a link between ethanol metabolism and the $\text{NAD}^+:\text{NADH}$ ratio (Chapter 2). Also, results from the preliminary M.Sc. metabolomics investigation on the effect of acute alcohol consumption on healthy subjects (Mason, 2010), on which my study was based, suggested that some of the observed metabolic changes were related to the

depletion of NAD⁺ (which results from the metabolism of the alcohol consumed). It was therefore proposed that an assessment of the effect of NAD (in the form of a commercial supplement) should be included in the experimental design of the present study.

Thus, an experimental design for an intervention study that could primarily serve as a well-designed follow-up of Mason's preliminary investigation (Mason, 2010), and which addresses the experimental aspects outlined above, was developed. This design was refined in order to make it feasible to carry out while still maintaining its capability to provide useful data.

In this case, an intervention study (also referred to as a case-control study) is defined as one where experimental subjects are selected from a certain predefined population, based on specific inclusion and exclusion criteria so that all subjects are comparable at the start of the study. Then, immediately after establishing the baseline, the total group is randomly split into a group that receives an intervention (such as an acute dose of alcohol for example) and a group that does not (control). The two groups are then followed up prospectively, and the changes in each group over time are measured. The differences between the two groups are then considered to be an indication of the effect of the intervention over time (Lesaffre & Verbeke, 2005).

A longitudinal study approach was used for the present study. This approach involves the repeated observation or monitoring of the same group of individuals over an extended period of time (following an experimental intervention, for instance, as was the case with this study). Data are first collected at the outset of the study, and may then be gathered repeatedly throughout its duration (Coggon *et al.*, 1997). Since longitudinal studies monitor the same group of individuals, the differences/changes observed in those individuals are more likely to be from the experimental intervention than other factors (Yee & Niemeier, 1996).

Another important factor to consider is the choice of biological samples. Urine, the biofluid used in my study, is frequently used in metabolomics studies for exploring systematic perturbations (Pasikanti *et al.*, 2008; Saude & Sykes, 2007; Saude *et al.*, 2007) since it:

1. is readily available, easily obtained, and less complex than other body fluids. Ease of collection also allows for multiple sampling, which greatly improves the reliability of statistical analyses and aids in monitoring disease progression and therapeutic response (Bernini *et al.*, 2011);
2. contains relatively few proteins and requires simpler sample preparation than other types of biofluids collected;
3. exhibits large variations in ionic strength, pH and osmolarity, especially under physiological stress;

4. is not under homeostatic regulation and therefore reflects metabolic dysregulation;
5. is an extra-cellular biofluid, which reflects a picture of metabolic activity at a specific time or over a certain period of time, and can be considered as a metabolic 'snapshot' or 'footprint' that integrates the metabolic responses of multi-cellular systems (Dunn *et al.*, 2008);
6. contains metabolites of many biochemical pathways, huge dynamic ranges of metabolic classes and concentrations, and an extreme diversity of both endogenous and exogenous (e.g. microbial, drugs, pollutants, industrial) chemicals, which were previously poorly characterized and understood, but gained improved understanding through the comprehensive and quantitative metabolome-wide characterisation of human urine (Bouatra *et al.*, 2013);
7. gives an averaged pattern of easily-excreted polar metabolites as a result of catabolic processes (Álvarez-Sánchez *et al.*, 2010; Want *et al.*, 2010).

It was agreed upon that the study would involve the hourly collection of urine samples for 6 hours from experimental subjects exposed to four interventions: (1) vehicle only, (2) vehicle plus an acute dose of alcohol, (3) vehicle plus an NAD supplement, and (4) vehicle plus an acute alcohol dose and NAD supplement. It was decided that data from 20 experimental subjects for each intervention were required for effective and useful statistical analysis and interpretation, but to use 24 subjects per intervention to account for outliers or for data of some subjects possibly needing to be removed from the study, e.g. due to partial participation or withdrawal.

Homogeneity of the experimental group is of critical importance in order to minimize non-relevant biological variation, especially in a GC–MS approach (Chapter 5), where variation due to external factors, such as diet, medication and lifestyle, are often also reflected. Therefore, this investigation was designed as a cross-over study, meaning that the same 24 experimental subjects were used in the four different interventions. Over the course of the study (proposed to be conducted on one day of the week for four weeks), the 24 subjects were randomly assigned to an intervention group until all 24 had participated once in each of the four interventions. This made the study group more homogeneous and eliminated the need for a large number of experimental subjects. Figure 4–1 represents the basic experimental design followed for my study.

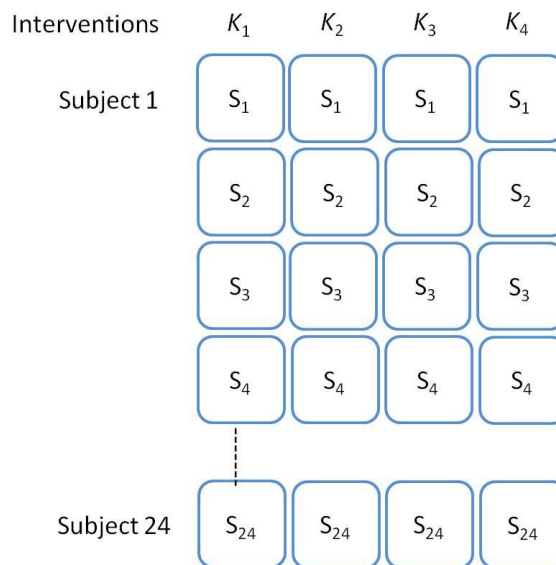


Figure 4-1: Structure of the basic experimental design for all experimental subjects. **K represents the interventions in the study.** Each of the 24 subjects was exposed to four interventions over the course of the study: K_1 : “vehicle only”; K_2 : “alcohol plus vehicle”; K_3 : “vehicle plus NAD”; K_4 : “vehicle plus alcohol and NAD”. For each subject, S_1 to S_{24} , urine samples were collected at six time points, as described.

In order to minimize consumption of various unwanted additives and other substances, it would have been preferable to use pure chemical ethanol for this investigation. However, for feasibility reasons, and due to the ethical constraints of using pure chemical ethanol, vodka was selected as the alcoholic substance to be used since it is quite palatable and contains relatively few additives compared to other liquors. From the literature (Numminen *et al.*, 2000 for example), it appeared that 1.5 g vodka per kg body mass was an appropriate acute dose of alcohol; 500 mL lemon-flavoured carbonated water was chosen to be used as the vehicle to dilute the vodka to be consumed. One capsule of NAD (50 mg) was selected as an appropriate dose for the exploratory study. Based on observations from a small orientation study, done in advance, it was decided that the NAD should be taken one hour before consumption of the alcohol in order to maximize its potential effect. The NAD used in my study was commercially available as a supplement (called *NAD assist*), marketed by Future Health (www.futurehealthsa.co.za). According to the description, the supplement is regarded as a formulation against chronic fatigue, and can be used for unlocking cellular energy potential, as well as to enhance well-being, the natural healing abilities of the body, mental clarity, alertness and concentration. It is further described as being used for its antioxidant and substance abuse benefits, including tobacco, alcohol, and drug abuse. Through personal communication, professionals at a clinic for alcoholism suggested that NAD could be a beneficial supplement to binge drinking individuals.

Based on this, NAD was included in the third and fourth interventions as an exploratory procedure, aimed at getting an indication of its potential effect on the metabolic profile following acute alcohol consumption. This exploratory study could pave the way for another appropriately designed experiment, aimed at elucidating the complete effect of commercial NAD supplementation on the effects of acute or chronic alcohol consumption, taking into account the biochemical individuality of the experimental subjects (Beger *et al.*, 2016).

In order to minimize experimental variation and the amount of non-relevant biological information obtained from the analyses, it was required that all experimental subjects refrained from the use of any supplementation, followed a similar dietary and lifestyle pattern, were in an overnight fasted state on the days of sample collection, and consumed only bottled water for the duration of the experiment and sample collection. For each of the four interventions, each of the 24 participants provided an early morning urine sample, collected one hour before the start of the experiment (time -1). Subsequent urine samples were collected just prior to consumption (time 0), as well as 1, 2, 3 and 4 hours thereafter. This gave a total of six samples per participant for each of the interventions. These samples were to be used for NMR (this chapter) as well as GC-MS (Chapter 5) analyses. In studies involving multiple analyses of the same sample it is advisable to sub-aliquot urine samples into appropriate volumes and sample containers in order to minimize the number of freeze-thaw cycles. Therefore, shortly after collection, each sample was divided into aliquots before being stored at -80°C in order to reduce the influence of freeze-thaw cycles. Each of the two approaches used in my study (NMR and GC-MS) required a method-specific measurement design, that was adapted from the general experimental design described above. The modified NMR measurement design is presented in the papers that emanated from this investigation (sections 4.4.2 and 4.5.2). The modified GC-MS measurement design is given in Chapter 5.

This experimental design resulted in the collection of a large number of samples to be analysed. This necessitated splitting samples across multiple analysis batches, where a batch is defined as a collection of consecutive injections belonging to different samples with no analytical replication (Kamleh *et al.*, 2012). In order to estimate any batch effect or other interfering analytical aspects, quality control (QC) samples — defined as a pooled sample of equal aliquots of all samples to be analysed — were included and analysed within each experimental batch. Thus, once all the samples from all the participants, interventions and times were collected, a 1 mL aliquot of each urine sample was thawed and combined to prepare this pooled QC sample, which was then divided into 15 mL aliquots and once again stored at -80°C. These QC samples were considered to be representative “mean” samples, theoretically containing all the analytes and metabolites that will be present in the experimental samples to be analysed (Gika *et al.*, 2007; Gika *et al.*, 2008; Kamleh *et al.*, 2012; t’Kindt *et al.*, 2009). The QC samples included in

each batch were prepared individually and in the same way as all the other experimental samples, and placed and analysed at regular intervals throughout the batches, as described in the separate NMR and GC–MS measurement designs.

4.3 Perspectives on NMR technology as a mode for the generation of metabolomics data

An essential aspect of metabolomics is the ability to generate biological information from the complex metabolite data obtained following analytical analyses of biological samples. NMR can be considered one of the most important analytical platforms for metabolomics studies (Chapter 5, Figure 5–1), owing to its inherent quantitative nature, and the wealth of chemical and spectroscopic information it can provide about the nuclei of organic molecules. However, the sensitivity of NMR analysis to the chemical environment adds certain analytical challenges — subtle differences in pH, ionic strength, temperature, and protein content between samples may result in differences in the NMR-detected peak positions and line widths for a given metabolite. In addition, a limitation of ^1H -NMR spectroscopy (used in this study) is the requirement for a distinct presence and orientation of protons in a compound. Thus, compounds without these chemical characteristics (that is, compounds without NMR-detectable protons) are said to be essentially “NMR invisible”, and will therefore be undetectable using ^1H -NMR. The application of NMR, and pattern recognition methods, in metabolomics studies have been extensively reviewed (Emwas *et al.*, 2015; Markley *et al.*, 2017; Smolinska *et al.*, 2012; Wishart, 2008b; Zhang *et al.*, 2010), and therefore no further, in-depth discussion is given here. However, given the importance of the experimental conditions for the generation of suitable NMR spectra, and the alternatives available to transform the spectral data into quantitative metabolite information, some key aspects of ^1H -NMR (as applicable to my investigation) will be discussed here.

4.3.1 Sample preparation

A major advantage of NMR over methods such as GC–MS, for example, is the fact that minimal sample preparation is required. Therefore, since little is changed in the sample itself before analysis, this method allows for a holistic view of the sample ‘as is’. Full details of the sample preparation procedure are given in Irwin *et al.*, 2016 (section 4.4.2 below), but, in short: (1) samples are centrifuged, which ensures the removal of interfering macromolecules in order to prevent peak broadening; and (2) a buffer, with a pH of 7 and containing a quantitative internal standard, is added to ensure a constant pH is maintained across all samples. This is an important requirement in NMR, and the first point of care regarding potential variation, since peaks are identified based on their chemical shift (positioning in the spectra relative to the peak of the internal standard), a characteristic which is pH sensitive.

4.3.2 ¹H-NMR analysis

The theory behind how NMR spectroscopy works is a complex, multi-faceted topic that could fuel many Ph.D. studies. NMR metabolomics, specifically, has been covered in various scholarly reviews (Emwas *et al.*, 2015; Markley *et al.*, 2017; Smolinska *et al.*, 2012; Wishart, 2008b; Zhang *et al.*, 2010), and will therefore not be repeated here. For the purpose of my study, however, it is appropriate to broadly outline the basic fundamental physics behind this technology.

First, it is important to mention that all charged nuclei possess a spin, and hence a magnetic dipole moment. In nature, the direction of this spin is completely random. Under an applied magnetic field, however, these nuclei align in either a north–south or south–north direction. While aligned under this applied magnetic field, a pulse of energy, in the form of a radio frequency (RF), is applied at precisely 90° to the nuclei. This pulse forces the nuclei into a higher energy state, and knocks the direction of the spin of the nuclei into a precessional orbit about the z-axis. The direction of the spin naturally returns to its lower energy state, with a subsequent loss of energy. This loss of energy, known as free induction decay (FID), undergoes Fourier transformation, which converts it from a time domain to a frequency domain, finally yielding a peak on a NMR spectrum.

While NMR is much more complex, and involves a great deal more detail than given here, this brief theory is important to note, since the ability to produce repeatable NMR spectra depends on the ability to maintain a constant aligned magnetic field, and an exact 90° RF pulse, and the ability to accurately record the FID. Herein lies the second critical analytical point of care and potential source of variation. In order to eliminate any potential variation, each sample analysed with NMR underwent a number of pre-calibration steps in the form of: (1) ‘locking and tuning’, to ensure that the probe (detector) correctly records the required FID data; (2) ‘shimming’, to align the magnetic field perfectly for each sample; and (3) calibrating the RF pulse to exactly 90° to the nuclei. Finally, the positioning of NMR peaks is also temperature sensitive, and therefore, during the entire experiment, a temperature control unit within the NMR probe maintained a constant temperature of 300 K. Thus, the implementation of these steps ensures that any technical variation caused by the NMR spectrometer is kept to an absolute minimum, thereby producing the most accurate and repeatable results.

4.3.3 Quality assurance

As described in my experimental design, the experimental samples were divided into batches, each batch containing the 24 samples (six samples for each of the four interventions) from a single subject. Also included in each batch, as part of the NMR measurement design, were

three pooled QC samples, prepared in the same way as the other samples, and loaded and analysed at the 1st, 14th and last (27th) position in each batch. One batch of samples was analysed per day for a period of approximately five weeks.

Collectively, this design yielded 72 QC samples across the 24 batches; 29 metabolites in these QC samples were constantly quantified and monitored throughout the experiment. The results, expressed as coefficients of variance percentages (CV%), are given as a bar graph in Figure 4–2.

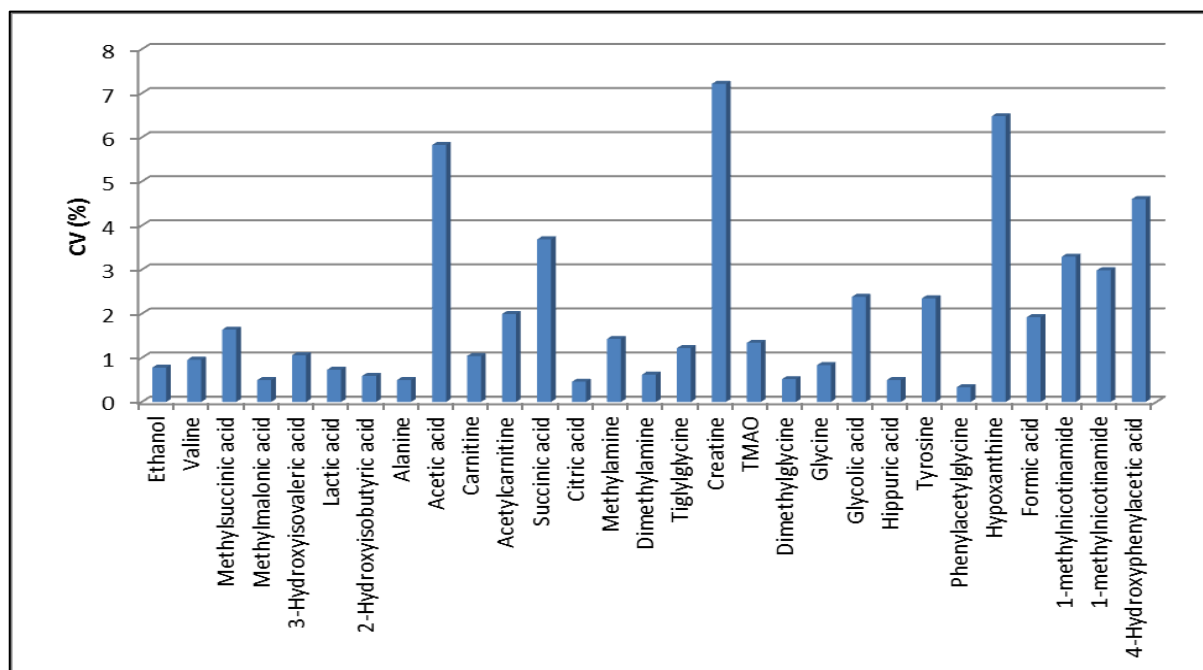


Figure 4-2: Bar graph illustrating the CV% values of the QC samples analysed using NMR spectroscopy.

As is evident from the CV% values in Figure 4–2, NMR is indeed a highly repeatable, and an analytically reliable, technique. The maximum CV% value was ~7% (for creatine), whereas ethanol — one of the most important metabolites quantified in my study — had a CV% value of less than 1%. These repeatability results unequivocally provide confidence in the ability of NMR to provide reliable quantitative results.

The remainder of this chapter describes the use of NMR as an analytical platform in determining the effect of consuming of the flavoured water vehicle, as well as in observing the metabolic perturbations that occur following acute alcohol consumption. The results from these investigations are presented as two peer-reviewed papers (sections 4.4.2 and 4.5.2, respectively). NMR was chosen to be used during both of these investigations owing to its high repeatability, as is evident from the highly repeatable results from the QC sample analyses

shown above. Additionally, due to its high volatility, urinary ethanol (a very important metabolite in this study) is not detectable by the GC–MS method used during my study. It can, however, easily be detected and quantified by means of NMR spectroscopy, making NMR a convenient and appropriate complementary analytical platform to use during this investigation.

4.4 An intervention study on the detoxification of benzoic acid

4.4.1 Background

The vehicle used for the alcohol and NAD interventions was aQuellé lemon-flavoured sparkling water (www.aquelle.co.za, a product of South Africa). It contained some fructose, vitamin C and citric acid as flavouring agents, sodium benzoate as preservative, and sodium cyclamate, aspartame and acesulfame K as sweeteners. Sodium benzoate, a known metabolite originating from the human gut flora, is excreted as hippuric acid, one of the dominant constituents of the human urine metabolome. A key requirement for insight into the effect of acute alcohol consumption was thus to determine *a priori* the effect, if any, of consumption of only the vehicle. This aspect of my project constituted the first of the four interventions, indicated as K_1 in Figure 4–1. The effects of the “vehicle only” intervention (as well as the effects of the three other interventions) were measured over a period of 5 hours, indicated in Figure 1 of Irwin *et al.*, 2016. This figure represents all the elements of the experimental design for the NMR analysis.

The outcome of the “vehicle only” intervention, presented in section 4.4.2 as a peer-reviewed paper (Irwin *et al.*, 2016), provided a number of important insights, including:

1. The value of the application of untargeted NMR to metabolomics data for the investigation of a single intervention.
2. A number of complementary approaches of data analysis that is required for the accurate analysis and interpretation of data collected from a longitudinal intervention study.
3. New insight into the metabolic pathways associated with benzoic acid detoxification.

4.4.2 Contribution towards a metabolite profile of the detoxification of benzoic acid through glycine conjugation: an intervention study (Irwin *et al.*, 2016)

RESEARCH ARTICLE

Contribution towards a Metabolite Profile of the Detoxification of Benzoic Acid through Glycine Conjugation: An Intervention Study

Cindy Irwin¹, Mari van Reenen^{1,2}, Shayne Mason¹, Lodewyk J. Mienie¹, Johan A. Westerhuis^{2,3}, Carolus J. Reinecke^{1*}

1 Centre for Human Metabolomics, Faculty of Natural Sciences, North-West University (Potchefstroom Campus), Potchefstroom, South Africa, **2** Department of Statistics, Faculty of Natural Sciences, North-West University (Potchefstroom Campus), Potchefstroom, South Africa, **3** Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands

☞ These authors contributed equally to this work.

* carools.reinecke@nwu.ac.za



OPEN ACCESS

Citation: Irwin C, van Reenen M, Mason S, Mienie LJ, Westerhuis JA, Reinecke CJ (2016) Contribution towards a Metabolite Profile of the Detoxification of Benzoic Acid through Glycine Conjugation: An Intervention Study. PLoS ONE 11 (12): e0167309. doi:10.1371/journal.pone.0167309

Editor: Viswanathan V Krishnan, California State University Fresno, UNITED STATES

Received: June 14, 2016

Accepted: November 13, 2016

Published: December 1, 2016

Copyright: © 2016 Irwin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are provided within the manuscript and the Supporting Information file [S2 File](#) (CI_unscaled raw vehicle data.xlsx).

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. CI received a post-graduate bursary from the National Research Foundation (NRF) of South Africa. The views and findings of this

Abstract

Benzoic acid is widely used as a preservative in food products and is detoxified in humans through glycine conjugation. Different viewpoints prevail on the physiological significance of the glycine conjugation reaction and concerns have been raised on potential public health consequences following uncontrolled benzoic acid ingestion. We performed a metabolomics study which used commercial benzoic acid containing flavored water as vehicle for designed interventions, and report here on the controlled consumption of the benzoic acid by 21 cases across 6 time points for a total of 126 time points. Metabolomics data from urinary samples analyzed by nuclear magnetic resonance spectroscopy were generated in a time-dependent cross-over study. We used ANOVA-simultaneous component analysis (ASCA), repeated measures analysis of variance (RM-ANOVA) and unfolded principal component analysis (unfolded PCA) to supplement conventional statistical methods to uncover fully the metabolic perturbations due to the xenobiotic intervention, encapsulated in the metabolomics tensor (three-dimensional matrices having cases, spectral areas and time as axes). Identification of the biologically important metabolites by the novel combination of statistical methods proved the power of this approach for metabolomics studies having complex data structures in general. The study disclosed a high degree of inter-individual variation in detoxification of the xenobiotic and revealed metabolic information, indicating that detoxification of benzoic acid through glycine conjugation to hippuric acid does not indicate glycine depletion, but is supplemented by ample glycine regeneration. The observations lend support to the view of maintenance of glycine homeostasis during detoxification. The study indicates also that time-dependent metabolomics investigations, using designed interventions, provide a way of interpreting the variation induced by the different factors of a designed experiment — an approach with potential to advance significantly our understanding of normal and pathophysiological perturbations of endogenous or exogenous origin.

investigation are those of the authors and do not reflect or represent any policies of the NRF or TIA funders. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Applications of metabolomics to intervention or challenge studies greatly enhance the holistic understanding of the effects of consumed substances on metabolic pathways [1, 2]. Data sets from intervention studies are, however, complex, as these investigations aspire to measure multiple metabolites in a biofluid, obtained from several experimental subjects, collected at different points in time and subjected to interventions from different consumed substances. In addition, these studies call for methods of data analysis specifically designed for longitudinal (time-dependent), multi-subject (data from several experimental participants), multi-group (intervention studies) and multivariate data [3, 4]. In this paper, we present the experimental design for an intervention study which includes the complex aspects mentioned above. The interventions were consumption of alcohol in the presence or absence of NAD, using flavored water as a vehicle. We generated matched-sample series through a cross-over study of participating subjects, collecting samples over a distinct time frame. The biochemical responses to the interventions were distinctly different: responses to alcohol and NAD intake resided in the intermediary metabolism, whereas those to exogenous substances in the vehicle involved detoxification through biotransformation mechanisms. Here, we present the full experimental design of the intervention study, but focus on the contribution from the biotransformation response by presenting the outcomes of vehicle consumption only. The results on the alcohol and NAD interventions will be published in a separate paper.

Benzoic acid was an important constituent in the vehicle used in the present study. Benzoic acid and its derivatives are routinely used as preservatives and flavoring agents in food products. Consequently, human exposure to them is quite common, and has raised concerns about potential public health consequences [5]. Evidence that benzoic acid is excreted as hippuric acid after enzymatic conjugation to glycine dates back to the 1950s [6], but different viewpoints seem to prevail on the physiological significance of the glycine conjugation reaction. Traditionally, glycine conjugation became part of the paradigm of detoxification, with the critical role of glycine conjugation for aromatic acids [7]. More recently, new views were proposed, shifting the focus to glycine homeostasis to assist in the regulation of body stores of glycine and other amino acids which are key neurotransmitters in the central nervous system (CNS) [8] or to serve as a molecular escort in the glycine deportation system to excrete excess glycine into urine as hippuric acid [9].

Detoxification pathways — the traditional viewpoint — can directly affect the integrity of multiple organs and hence can be widely involved in a variety of human conditions, such as health [10], cometabolism in humans with the gut microbiome [11], disease therapy [12] and aging [13]. Lipophilic endogenous or exogenous xenobiotics are first metabolized by the phase 1 detoxification system, which converts the compounds into substances having a hydrophilic functional group for increased solubility. The phase 2 detoxification system involves conjugation reactions [14]. The phase 2 system comprises several enzymes, with glycine-N-acyltransferase (GLYAT) being the key enzyme in glycine conjugation.

Invoking the roles of aromatic acids and gut metabolites to regulate blood levels of glycine — the new viewpoint — relates to that of glycine as neurotransmitter in the CNS. A putative role for it as a neurotransmitter dates back to the observation that the concentration of glycine in spinal cord tissue is far higher than elsewhere in the brain [15]. The intracellular concentration of glycine is regulated mainly by a glycine transporter, acting in synergy with glutamic acid decarboxylase [16]. The proposed glycine deportation system, moreover, functions as a homeostatic regulator of the GLY central pool and, by definition, also the GLY tissue pool [9].

GLYAT (reviewed by Badenhorst *et al.* [17]) is located in the mitochondria of mammalian liver and kidney and is a member of the superfamily of N-acyltransferases. It has been

demonstrated that several GLYAT species [glycine N-acyltransferase (EC 2.3.1.13), glutamine N-acyltransferase (EC 2.3.1.68) and glycine N-benzoyltransferase (EC 2.3.1.71)] are closely related but distinct mitochondrial enzymes [18], although originally regarded as being a single enzyme [19]. Benzoyl-CoA, salicyl-CoA, and certain short-, straight- and branched-chain fatty acyl-CoA esters are substrates for the former enzyme, and glycine is the acyl acceptor for all these enzymes. As glycine and glutamic acid are known to cross the blood–brain barrier freely, it is assumed that removal of GLY from the GLY central pool will lower levels of GLY in the CNS [9].

Commercial flavored water, containing benzoic acid (the preservative and co-substrate in glycine conjugation), was used as vehicle for the present intervention study, which opened the opportunity for a metabolomics investigation on controlled benzoic acid consumption. The intervention study consisted of 24 experimental cases (healthy males between 20 and 24 years of age), and used a metabolomics approach for the experimental design and data analysis. The study included four treatments and six urine samples per experimental case were collected over a period of five hours and used as biofluid for the metabolomics measurements. We selected proton nuclear magnetic resonance ($^1\text{H-NMR}$) spectroscopy as the technology for the generation of the metabolomics data as it is the recognized method of choice for an untargeted coverage of the metabolome. The outcomes offered several intriguing views on the metabolites formed in each of these interventions. The results using the vehicle as control provided novel insights into the metabolite profile following benzoic acid consumption.

Time-dependent intervention studies require more technical statistical methods for data analysis, forming another central aspect of this paper. Metabolomics data are mostly represented as a matrix of controls and cases (rows) measured over the metabolites that reflect the perturbation under investigation (columns). Such data are traditionally processed through multivariate statistical methods, using mostly unsupervised principal component analysis (PCA) and a supervised method such as partial least-squares discriminant analysis (PLS-DA), to identify metabolites that differ significantly between the test groups studied [20]. In human intervention studies, individual variation tends to dominate the often less pronounced metabolic perturbations due to the intervention. A reasonably large number of cases is mostly required as well as a sufficient number of time points to cover the effect of the intervention. We used a cross-over experimental design, which generated several hundred samples which requires a different approach from the traditional multivariate methods for processing the data and a more extensive validation of the results than for standard metabolomics studies.

For optimal insight into the intervention involving vehicle consumption, we followed a novel approach of complementary statistical methods: (1) the traditional multivariate and univariate methods of analysis were included as the first approximation of the biological profile following the intervention; (2) ASCA (ANOVA-simultaneous component analysis), developed for analyzing designed metabolomics data [4], was used next as a method that deals with multivariate data sets based on a defined experimental design, including time-dependent data; (3) RM ANOVA (repeated measures analysis of variance), which is similar to ASCA as it accounts for the experimental design of the data (although RM ANOVA was used to assess each metabolite individually to find those metabolites that changed significantly with time); finally, (4) unfolded PCA [21] was applied to gain insight into the global (i.e. multivariate) effect of the vehicle over time. The combination of these complementary statistical methods and transdisciplinary approach of metabolomics to the research lead to a more holistic view of the data. This proved essential in elucidating the effect of the intervention and provided novel insights into benzoic acid biotransformation, not observed by more reductionist methods of analysis.

Materials and Methods

Chemicals and reagents

The substances used for the intervention reported here were flavored water (aQuellé lemon-flavored sparkling water: fructose and citric acid flavoring; sodium benzoate preservative; sodium cyclamate, aspartame, acesulfame K sweeteners; vitamin C – www.aquelle.co.za, product of South Africa), commercial water (Valpré still spring water, inorganic contents specified – product of South Africa). The internal standard for the ^1H -NMR analysis was trimethyl-2,2,3,3-tetradeuteriopropionic acid (TSP, sodium salt; Sigma Aldrich).

Experimental design

The intervention study was based on an observation in a preliminary metabolomics investigation on the effect of acute alcohol consumption in healthy cases [22] that indicated some NAD depletion. The experimental group consisted of 24 medically confirmed healthy males, between 20 and 24 years of age, residing in the same student hostel of North-West University (South Africa). They were neither alcohol addicts nor total abstainers, but confirmed their use of alcohol on a moderate social level. No cases took any medication, all were asked to refrain from vitamins, minerals, and other supplementation, and were requested to follow a similar dietary and lifestyle pattern for the duration of the study. The protocol was approved by the Health Sciences Ethical Committee of North-West University (Ethical approval number: NWU-00045-12-S1), conducted in accordance with guidelines for good clinical practice and performed at the Health Clinic of the university. A medical doctor and nurse were present during the entire period of intervention and all cases could leave the premises only after approval by the doctor.

The experiments were conducted on Saturday mornings between 08:00 and 12:00. All subjects had to abstain from breakfast and had to provide an early morning urine sample, collected one hour before the start of the experiment (time –1). The treatments consisted of a measurement of the baseline effect of imbibing the vehicle (500 mL flavored water) that was also used in the remaining three interventions. All the cases were randomly assigned to a treatment/intervention group until all 24 participated in all four interventions. All cases were provided with 1.5 L pure spring water, which was the only substance that could be taken over the period of the experiment. High diurnal variation in urine is well established, and a validated dietary exposure biomarker discovery protocol demonstrated that top-ranked signals discriminating between fasting and 2–4 hour postprandial urine samples could be linked to metabolites abundant between components of a standardized dietary intervention [23]. Thus, urine samples were collected at time zero, just prior to consumption, followed by four further samples at 1, 2, 3 and 4 hours after imbibing the vehicle, providing six samples in total for each case. Owing to commitments of some participants, the experiments were performed over a period of 7 consecutive Saturdays, and all samples were treated and stored as described below.

Measurement design

Because of the large number of samples produced by this study, their analysis spanned a long time, which necessitated splitting samples across multiple analysis blocks or batches. Analytical assessment of metabolomics data generation proved the highly repeatable nature of NMR data measurements [24]. Although repeatability and reproducibility are not concerns in NMR analyses, the measurement design included the use of pooled quality control (QC) samples to estimate any batch effect or other interfering analytical aspect. Each batch contained the 24 samples (corresponding to 4 treatments across 6 time points) of a single case and three QC samples.

The 27 samples from each batch were analyzed in the following order:

QC₁ [S₋₁S₀S₁S₂S₃S₄]_{Vehicle} [S₋₁S₀S₁S₂S₃S₄]_{Alcohol} QC₂ [S₋₁S₀S₁S₂S₃S₄]_{NAD}
[S₋₁S₀S₁S₂S₃ S₄]_{NAD+Alcohol} QC₃

where S₋₁ represents the sample collected at time -1, S₀ represents the sample collected at time 0, and so on.

Sample collection, characterization and storage. Six samples per intervention were collected from each student. This gave a total of 24 urine samples from each case over the course of the study. One 5 mL and two 1 mL vials were used to provide aliquots of each of the urine samples; these aliquots, together with the remainder of the bulk urine samples, were stored at -80°C. Once all the urine samples were collected, one 1 mL aliquot of each was thawed and combined to prepare a QC sample for the experiment as a whole. This QC sample was then divided into 15 mL aliquots and once again stored at -80°C.

Sample preparation and ¹H-NMR analysis. Spectral analyses were conducted at the NMR facility of the Centre for Human Metabolomics at North-West University. Prior to analysis, an aqueous 1.5 M KH₂PO₄ deuterated buffer solution at pH 7.4 was prepared [25]. This solution served to lock the signal during analysis, maintained a stable pH in the sample and contained TSP as the internal standard to provide a chemical shift reference of δ = 0.00. The urine samples, stored at -80°C, were thawed at room temperature for analysis. A 600 μL volume of each sample was centrifuged at 12 000 × g for 5 min to remove any sediments or debris. A 60 μL volume of buffer solution was added to 540 μL of the supernatant, vortexed and transferred to a 5-mm NMR tube.

Each sample so prepared was analyzed on a Bruker Avance III HD 500 MHz NMR spectrometer equipped with a triple-resonance inverse (TXI) ¹H{¹⁵N, ¹³C} probe head and x, y, z gradient coils. ¹H spectra were acquired as 128 transients in 32K data points with a spectral width of 6002 Hz. The sample temperature was maintained at 300 K and the H₂O resonance was pre-saturated by single-frequency irradiation during a relaxation delay of 4 s, with a 90° excitation pulse fixed at 8 μs. Shimming of the sample was performed automatically on the deuterium signal. The resonance line widths for TSP and metabolites were <1 Hz (measurements at half the height of the peak). Fourier transformation and phase and baseline correction were done automatically. The software used was Bruker Topspin (V3.2) and Bruker AMIX (V3.9.9) [26].

All urine samples were normalized with reference to the creatinine CH₂ peak at 4.05 ppm. We employed two methods of spectral analysis: (1) the first method consisted of equidistant binning, using a bin width of 0.02 ppm applied to the spectral region of 0.5–10 ppm (344 bins). This gave a total of 467 integrated units per NMR spectrum, excluding the water region, for each sample for multivariate analysis (S2 File). Based on previous empirical experience with NMR spectral analysis, we defined a threshold value of 2 × 10⁶, being approximately the limit of detection of metabolomic substances presumed to be present in a spectral bin. All values below this threshold were set to zero.

The methodology of processing NMR spectra is well known. Powers [27] describes two schools of thought in this method: peak alignment and binning, the latter of which we used in this study. The equal-binning procedure masks subtle chemical shift differences and hides potentially significant changes of low-intensity peaks, but incurs the risk of splitting peaks or spectral features between bins. The second method used quantified metabolites, derived from variably-sized binning of the original spectral peaks above the noise level. This alternative method prevents peak division between multiple bins, avoiding the problems incurred in the first method. This second approach was specifically applied for the accurate identification and quantification (μmol/mmol creatinine) of discernible metabolites, generating data for univariate analysis. In addition, we used both the 1D and 2D J-resolved (JRES) approach to further characterize guanidinoacetic acid observed in some urine samples (Section 4 and Figure J in S1 File).

Data analysis

A tensor (Fig 1) was used for three modes of statistical analyses: (1) cross-sectional analysis of time points using traditional multivariate (PCA and PLS-DA) as well as univariate analyses — Wilcoxon signed rank test (WC p-value) and associated fold changes (FC), generated for each bin; (2) ASCA performed across all times; and (3) univariate RM ANOVA for each bin. Further detail on these methods is presented in [S1 File](#).

The original spectral data, derived from the cases following the vehicle intervention, sampling at six time points and including the QCs (Section 1 and Table A in [S1 File](#)), yielded a total of 144 samples, as shown in the flow diagram (Fig 2) and illustrated through a representative NMR spectrum for one QC sample (Fig 3). However, during the experiment one participant did not complete all four interventions, thus yielding 138 samples obtained from the remaining 23 cases, which were used in the data analysis. The original spectral data were pre-processed by normalization relative to the creatinine content (between 4.05 and 4.07 ppm); replacing very low values with zero and performing a 50% zero-filter. Case reduction was based on batch comparisons of the QC samples (detail presented in [S1 File](#)) — two batches were found to be outliers and were removed.

A list of important bins was generated by each of the three modes and combined using a Venn diagram approach to identify a final shortlist of important spectral bins. This shortlist contained information on the main metabolites that reflected the intervention.

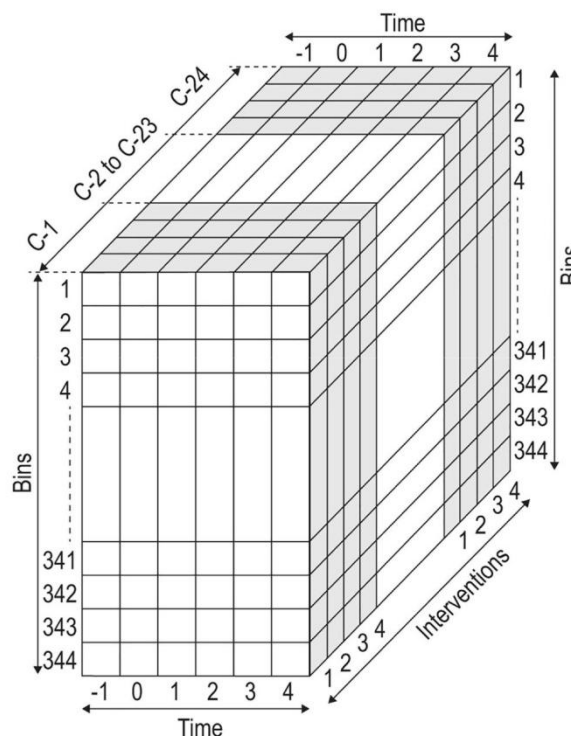


Fig 1. Representation of all elements of the experimental design. The tensor for the complete design consisted of four dimensions: the experimental cases (total 24 — the shaded areas in the figure are for cases 1 and 24), time points (6, hourly), interventions (4, including the vehicle only) and spectral bins (344, following data pre-treatment). The total number of 198 144 data points ($= 24 \times 6 \times 4 \times 344$) thus required bioinformatics analysis to uncover the information from the four interventions. All 24 experimental subjects agreed to participate in each of the four interventions (i.e. a cross-over design), indicated by 1 (vehicle only), 2, 3 and 4. For each of the 24 cases, urine samples (from which the NMR spectral bins were generated) were collected at one hour prior to the intervention (time -1), just before the intervention (time 0) and then at hourly intervals for 4 hours (times 1–4). The results reported in this paper apply to only one of the treatments (consumption of flavored water), which yielded 49 536 data points.

doi:10.1371/journal.pone.0167309.g001

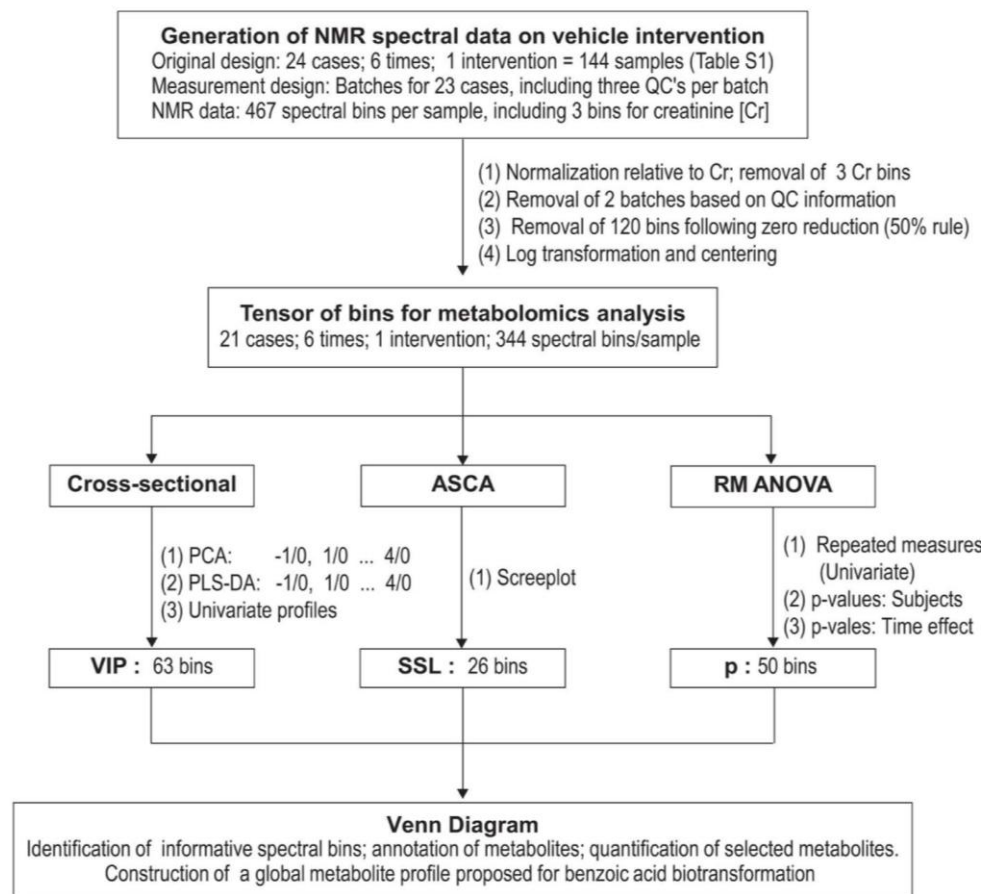


Fig 2. Flow diagram indicating the main lines of activity following data generation, identification and quantification of important metabolites on the intervention towards the proposed biological interpretation. VIP refers to variables important in projection, based on a PLS-DA; SSL refers to the sum of the squared loadings of the ASCA model; p refers to the RM-ANOVA p-values for the time effect.

doi:10.1371/journal.pone.0167309.g002

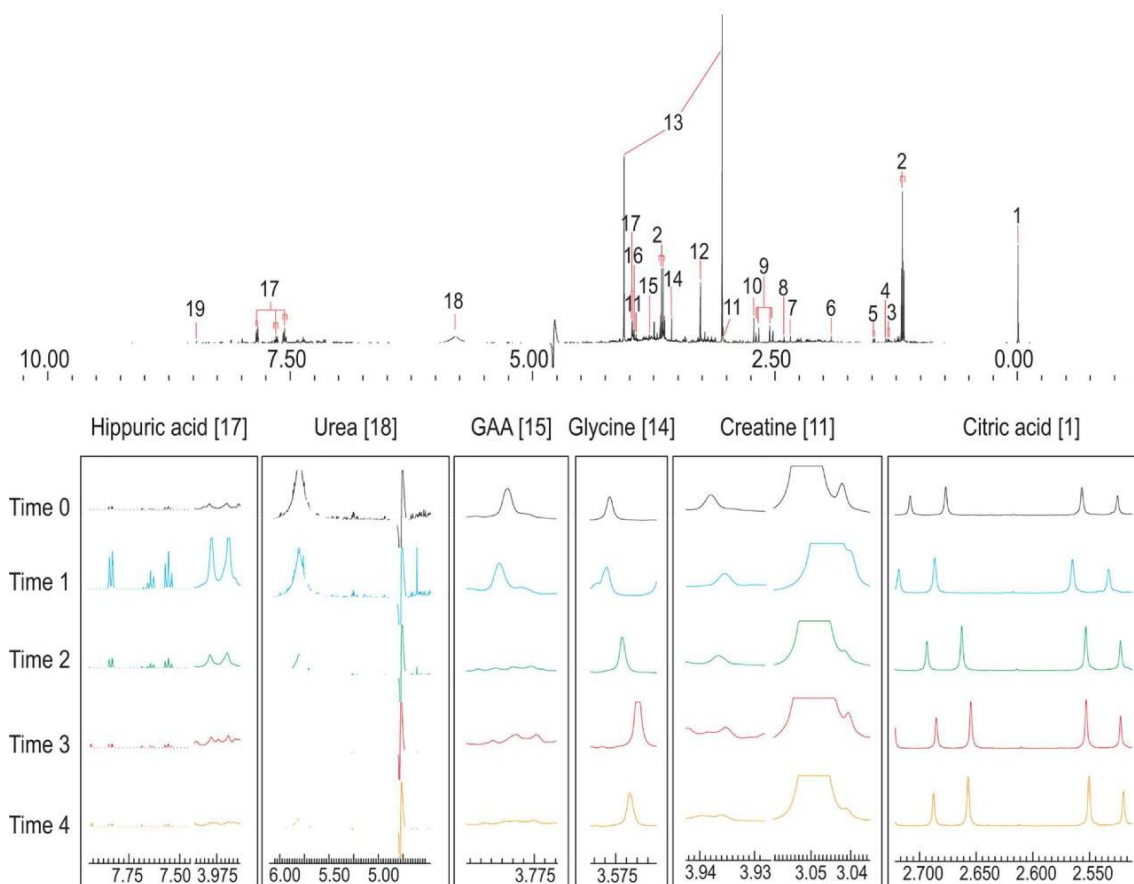


Fig 3. 500 MHz ^1H -NMR spectra of urine. Top spectrum taken from QC sample; numbers indicate the following metabolites: 1: TSP; 2: ethanol; 3: lactic acid; 4: 2-hydroxyisobutyric acid; 5: alanine; 6: acetic acid; 7: pyruvic acid; 8: succinic acid; 9: citric acid; 10: dimethylamine; 11: creatine; 12: trimethylamine-N-oxide (TMAO); 13: creatinine; 14: glycine; 15: guanidinoacetic acid (GAA); 16: glycolic acid; 17: hippuric acid; 18: urea; 19: formic acid. The six important metabolites (high VIP values) are highlighted below to qualitatively show differences over time 0 (black), 1 (blue), 2 (green), 3 (red) and 4 (orange) — scaled according to creatinine. Apart from citric acid and hippuric acid, the detoxification product of benzoic acid, no other constituents from the vehicle appeared to be detectable in the NMR analysis.

doi:10.1371/journal.pone.0167309.g003

Results

Cross-sectional analysis

One student did not complete the full intervention program, and was thus removed from the group. All results were finally based on the 344 ^1H -NMR profiled bins for 21 cases, following the elimination of two outlier batches (i.e. data from two cases), across the 6 time points. General time-dependent changes in the spectral data were first evaluated with regard to the baseline, i.e. time 0. Cross-sectional comparisons of time points –1, 1, 2, 3 and 4 hours vs. time 0 were performed using three multivariate approaches (unsupervised Euclidian and Ward hierarchical cluster analyses presented as dendrograms, unsupervised PCA and supervised PLS-DA models) and the combination of two univariate approaches (the p-values from Wilcoxon signed rank test and fold change (FC) values presented as Volcano plots). The data were log transformed (shifted natural logarithm with shift parameter set to 1) and centered prior to performing PCA and PLS-DA. The remaining methods were applied to untransformed data. The statistical techniques used and the results of all time points for the cluster analyses, Volcano plots, as well as for the PCA and PLS-DA, are discussed in detail in [S1 File](#) and Matlab coding given as supplementary file ([S3 File](#)).

[Fig 4](#) indicates the data from time –1 hour, 1 hour and 4 hours relative to time 0 for the cluster analysis and Volcano plots.

[Fig 4A](#) shows little separation (time –1 relative to time 0) between the individual cases, indicating a perceived closeness of spectral data encapsulated in the NMR bins for the individual cases. The metabolite profiles detected in the urine of most samples collected immediately prior to consumption of the vehicle thus represent the profile following 12 hours fasting. This conclusion also holds for the case indicated as an outlier bin within the group as a whole (indicated by a red dot in [Fig 4D](#)). The distinct group separation after vehicle consumption ([Fig 4B and 4C](#)) clearly indicates a metabolic perturbation. The results for the other two multivariate methods (PCA: Figure D and PLS-DA: Figure E in [S1 File](#)) confirmed the main observations shown in the dendrograms: no group separation between the data obtained at time zero and one hour before the intervention (Figures D(A) and E(A) in [S1 File](#)), clear separation one hour after consumption, i.e. time 0 vs. time 1 (Figures D(B) and E(B) in [S1 File](#)). Characteristic of multivariate methods, the group separations were more evident when applying a supervised (PLS-DA) as opposed to an unsupervised method (PCA). The Volcano plots revealed that a single bin indicated a significant ($p \leq 0.05$) up-regulated value ($FC \geq 2.0$) at time 0 with respect to time –1 ([Fig 4D](#)), which increased to a total of 31 up- and down-regulated values ([Fig 4E](#)) one hour after the intervention ($p \leq 0.05$ and $|FC| \geq 2$) and becoming abundant after 4 hours ([Fig 4F](#)), as compared to the baseline measurement (time 0).

Inter-individual variation following vehicle consumption

[Fig 5](#) represents the scores based on the first two principal components (PC1 and PC2) of the PCA model for the unfolded data tensor. The data tensor was unfolded in time as described and illustrated in [S1 File \[21\]](#). The unfolding transforms a three-dimensional tensor into a two-dimensional matrix and thus allows for PCA.

PCA of the unfolded tensor provides insight into the effect of the vehicle in time on the bins (indicated by the ellipses and centroids in [Fig 5](#)) as well as for individual cases, shown by the dots and trajectories in [Fig 5](#). It is evident from [Fig 5](#) that there is a change in the global profile (i.e. across all cases) over time. Times –1, 0, 2, 3 and 4 appear to be quite similar, with a dramatic change visible after 1 hour (time 1). It was evident from the prior analyses (Figs 4 and 5) that the urinary spectral profiles following the consumption of flavored water suggested notable changes in the metabolite profiles of all individuals in response to the intervention. To illustrate the inter-individual variation, the centroids of the PCA for the group (all 21 cases) and individual cases were tracked and highlighted over the period from time 0 to time 4 hours ([Fig 5](#)). The trajectories drawn in [Fig 5](#) represent only three selected individuals, for clarity. The selected trajectories (as well as for the group as a whole, [Fig 5](#)) indicate clear similarities in the individual responses to vehicle consumption in time: the profiles of the trajectory from time 0 to 4 hours were comparable in some respects. However, distinct differences were also noted: the response at time 1 hour following vehicle consumption in case 9 gives the impression of being an outlier. However, since the QC samples for this individual (linked to a given batch) were not outliers, we attribute this variation to the unique response of the case to the vehicle. Since we are interested in these kinds of responses, the observation was retained. Overall, the observations shown in [Fig 5](#) indicate that the cases themselves are a noteworthy source of variation.

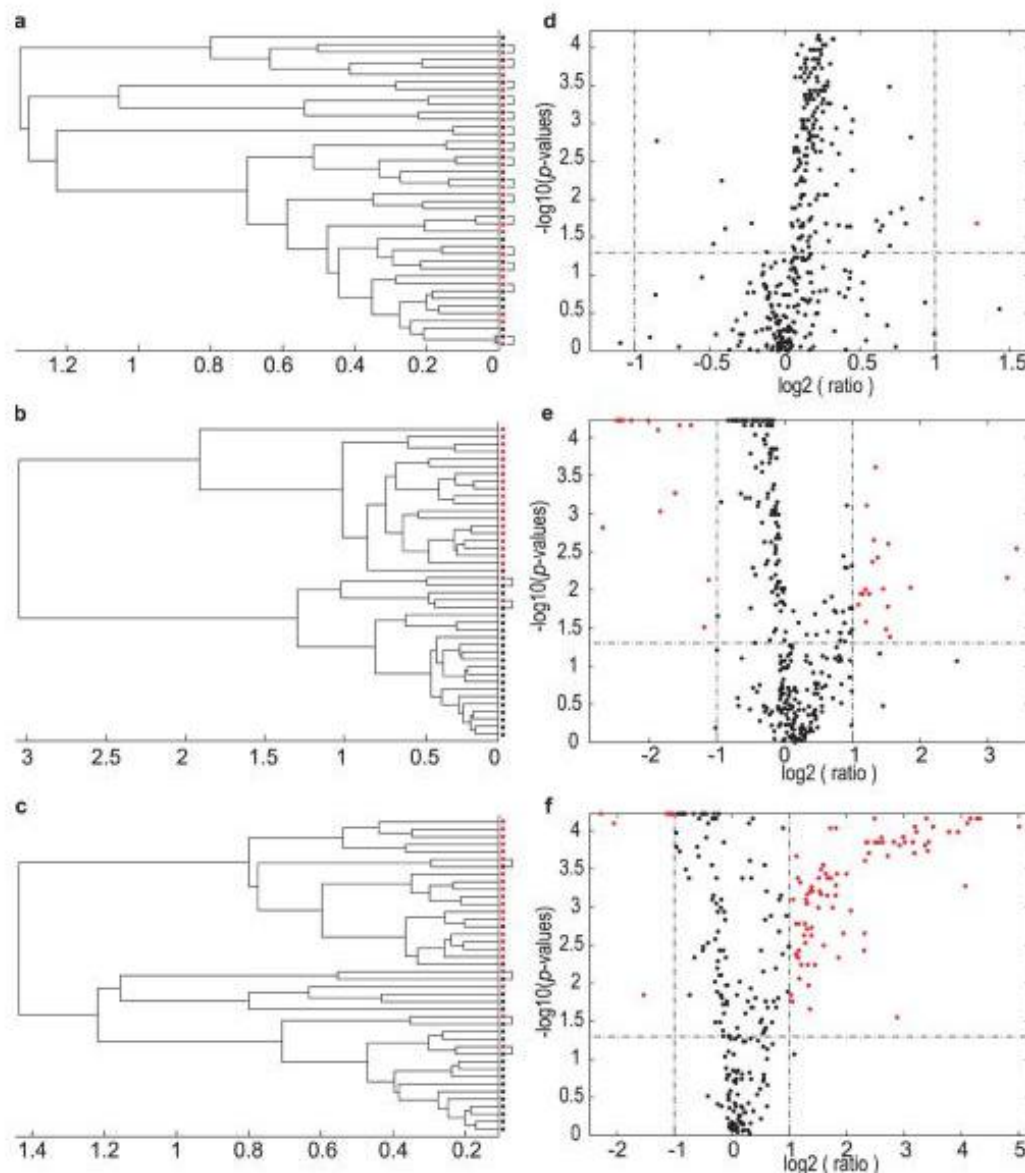


Fig 4. Group separation among experimental groups through dendrograms and Volcano plots based on equidistant binning spectral data. (a–c) Dendrograms from cluster analysis of 21 cases are shown. The dendrograms were constructed based on subsets of the data representing time 0 along with the data obtained one hour before the intervention (time –1); one hour; and four hours following the intervention (times 1 and 4), respectively. Data for all cases at time 0 are shown as black dots and at the other time slots as red dots. Time points from the same individuals that clustered close together are indicated by squared brackets, linking them. (d–f) Volcano plots for the same time points as in a–c indicate the distribution of the individual bins, based on FC and WC p-values. Demarcation of important bins is shown by the horizontal and vertical dotted lines. The number of influential bins is shown as red dots in the rectangles of the upper left and right segments of each Volcano plot.

doi:10.1371/journal.pone.0167309.g004

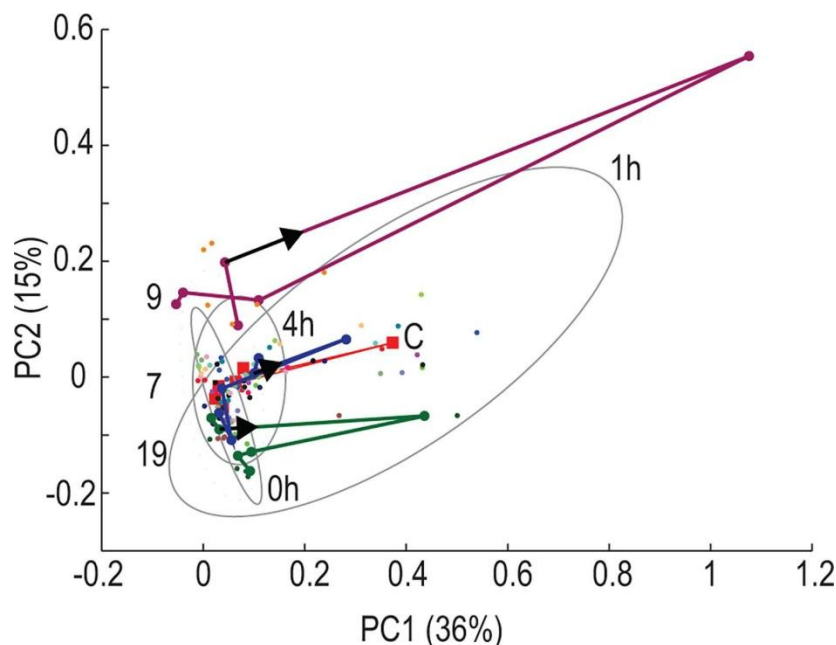


Fig 5. Unfolded PCA Scores Plots. PC1 and PC2 for all individual cases (colored dots) and indication of the 95% confidence ellipsoids for scores of PC1 and PC2 for times -1 , 1 and 4 hours. The centroids for these three clusters are indicated as red squares. The trajectories of the spectral profiles of three individuals are also shown. The trajectories of three cases (numbers 9 , 7 and 19 are shown as purple, blue and green lines, respectively) illustrate the inter-individual responses to the intervention. The direction of the trajectories is indicated by the short black arrow, starting from time 0 , for clarity.

doi:10.1371/journal.pone.0167309.g005

Longitudinal response to vehicle consumption

The results presented thus far suggest that the variation in the total data set was based on the intervention and superimposed by the normal inter-individual heterogeneity and time effects and by their interactions.

For optimal insight into the effect of the intervention, and specifically to identify the most important metabolites reflecting the metabolic perturbation induced by the flavored water, we used three complementary approaches. The traditional PLS-DA (illustrated in Figure E in [S1 File](#)) was used to find an approximation of the biological profile following the intervention. We identified 63 bins as important based on a maximum VIP over the five comparisons greater than 2. Second, ASCA ([Fig 6](#)) was applied as it provided an approach which was developed to deal with multivariate data sets based on a defined experimental design, including time-dependent data [4]. Again, the data were log transformed and centered. Similar to the unfolded PCA result, the plot of the sum of the ASCA effects and projected residuals onto the first two components of the effects subspace [28], shown in [Fig 6](#), indicate that observations at times -1 , 0 are quite similar, indicating a state of homeostasis that existed after the fasting period. A marked change became visible one hour after the intervention. This was followed by comparable profiles, especially between times 3 and 4 , indicating a return to a state of homeostasis, which was not identical to the state prior to the intervention, as will be discussed below. The sum of the squared loadings (SSL) of the ASCA model were ranked in decreasing order and used as a scree plot (Figure G in [S1 File](#)) to identify 26 bins with notably higher SSL values.

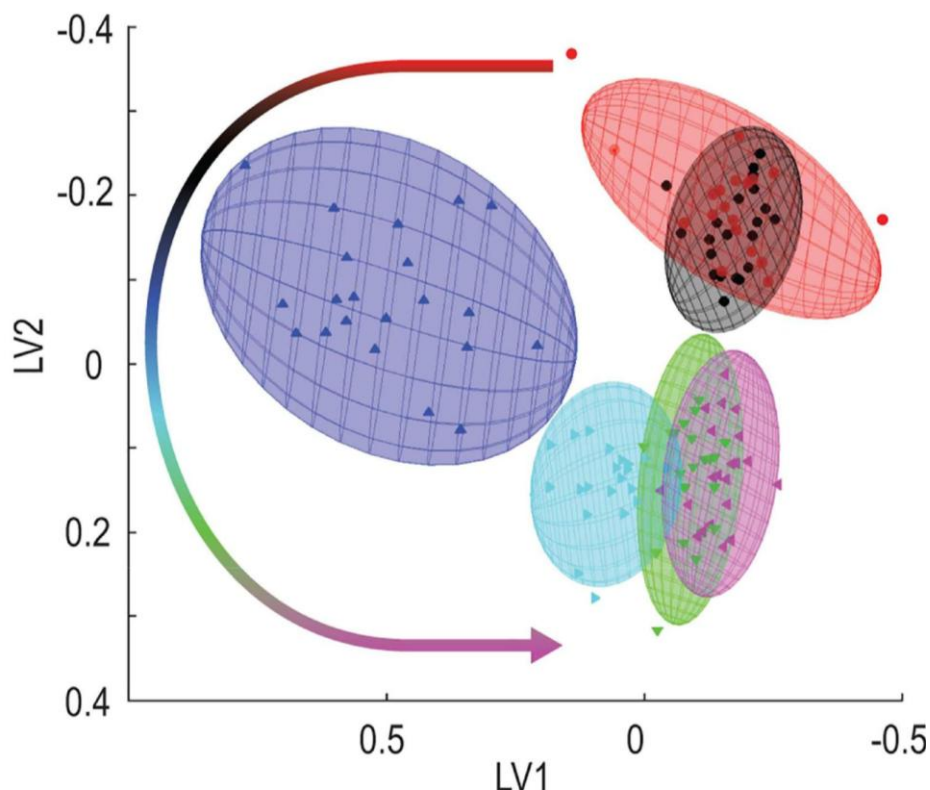


Fig 6. Application of ASCA to the 344 NMR spectral bins for 21 individuals over the period of the intervention. Plots of the sum of the ASCA effects and projected residuals are color coded according to time following the intervention (−1, 0, 1, 2, 3 and 4 hours shown in red, black, blue, light blue, green and pink, respectively), with the arrow showing the time-dependent trend, using the same discriminating color sequence.

doi:10.1371/journal.pone.0167309.g006

Third, RM ANOVA was used to assess each bin individually to identify bins that changed significantly in time, from which we selected the top 50 bins (p-values less than 0.00001) which appeared to be most informative. Data, generated from complex experimental designs such as the intervention study presented here, produce information that intersect in various ways.

Identification of important variables

A Venn diagram (Fig 7) was used to visualize the counts (number of bins) of the lists of important bins sharing some properties. Three properties were selected: maximum VIP 2 from the PLS-DA, SSL > 0.01 from the ASCA, and p-values from the RM ANOVA as defined above. A total of 29 bins were shortlisted, derived from being present in at least two of the three lists and which were unique in the intersection reflected by the shaded areas in Fig 7. It is interesting to note the power of the ASCA approach as it was able to identify 26 of the 29 bins, but this does not imply that variables not selected (e.g. the 29 RM ANOVA metabolites) are all unimportant as each method selects variables in its own way. For our purpose we concentrated, however, on the shared metabolites; how this was achieved is described in more detail in Section 3.6 of S1 File.

The list of 29 bins provided the final selection of bins to be identified as important metabolites and quantified to concentrations. Detailed information on the application of ASCA, RM ANOVA and unfolded PCA is presented in Section 3 of S1 File.

Analysis of the spectral characteristics of the 29 bins indicated that 23 bins related to endogenous human metabolites. Spectral information from four bins could not be assigned to any known chemical substance and two bins indicated exogenous contaminants.

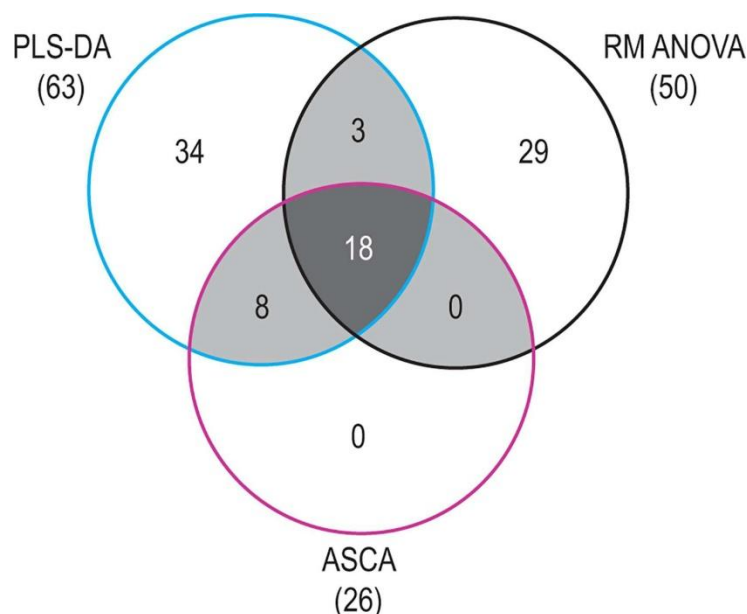


Fig 7. Venn diagram. Bins which are common to any two statistical approaches are indicated by the lightly shaded areas, and the 18 bins that are common to all three are shown in the central, heavily shaded area.

doi:10.1371/journal.pone.0167309.g007

Some bins (indicated in brackets) had structural information on the same metabolite, which resulted in a list of six key metabolites: urea (10 bins), hippuric acid (8 bins) and citric acid (2 bins) as well as creatine, 3-methylhistidine and guanidinoacetic acid each with one bin. One bin for each of the six metabolites, as well as for glycine (not identified as a highly perturbed metabolite), were quantified for all time points, as summarized in [Table 1](#).

Excretion kinetics for all six these metabolites are shown in [S1 File](#) (Figures H(A)–H(F)). The most conspicuous change — indicating the highly efficient detoxification of benzoic acid — occurred in the more than fourfold ($p < 0.0005$) increased urinary excretion of hippuric acid within one hour of consuming the flavored water, and its return to the same level as before the intervention ([Table 1](#) and Figure H(A) in [S1 File](#)). As indicated by the unfolded PCA scores ([Fig 5](#)), the trajectories of individual cases (illustrated for cases 7, 9 and 19) tended to return to positions close, but not identical, to the pre-intervention positions. The excretion kinetics of the six metabolites (Figure H in [S1 File](#)) and the Wilcoxon signed rank p-values ([Table 1](#)) gave strong numerical indications for this observation: (1) Hippuric acid dominated the profiles, and returned to practically the same state as prior to the intervention ($p = 0.689$, Time 0 vs Time 4). (2) The levels of guanidinoacetic acid ($p = 0.028$) and creatine (0.003) were significantly lower than prior to the intervention (Time 0 vs Time 4). (3) The excretion of citric acid ($p = 0.002$) and urea ($p < 0.0005$) were still after four hours significantly higher than prior to the intervention (Time 0 vs Time 4). Moreover, the excretion profiles of some unknown and unimportant metabolites, gave further support for the observed trajectory profiles (Figure I in [S1 File](#)).

The interpretation of changes in the excretion of these remaining metabolites paved the way for the construction of a metabolic profile, reflecting the consequences of detoxification of a single xenobiotic — benzoic acid as used here — as discussed below.

Table 1. Quantified data on metabolically important metabolites. All quantified values included are from NMR-determined urine analyses; the references are all from the Human Metabolome Database, and are expressed as μ moles metabolite/mmol creatinine. WC p-values are based on the comparison of the respective metabolite concentrations from time = 1 hour to 4 hours, relative to time zero.

Information on important metabolites		Time-dependent characteristics of important metabolites following flavored water consumption								
Metabolite	Normal values	Mean	Mean	WC	Mean	WC	Mean	WC	Mean	WC
(δ [ppm], multiplicity and chemical group)	Mean [Range] [Reference]	[SD] Time:0	[SD] Time:1	p-value	[SD] Time:2	p-value	[SD] Time:3	p-value	[SD] Time: 4	p-value
Citric acid (2.59 AB δ [(CH ₂) ₂])	203[49–600] [35]	213 [115]	264 [137]	<0.0005	322 [139]	<0.0005	320 [126]	0.001	306 [117]	0.002
Hippuric acid (3.97d - [CH ₂])	217[28–610] [35]	210 [100]	884 [268]	<0.0005	449 [170]	<0.0005	285 [154]	0.021	228 [142]	0.689
Glycine (3.57 s δ [CH ₂])	106[44–300] [35]	94 [63]	114 [58]	0.001	117 [37]	0.005	115 [34]	0.014	108 [32]	0.122
Guanidinoacetic acid (3.78 s δ [CH ₂])	89[11–124] [36]	97 [193]	77 [200]	0.173	0 [0]	0.028	0 [0]	0.028	0 [0]	0.028
Urea (5.78 broad δ [(NH ₂) ₂])	12285[174–49097] [35]	479 [230]	751 [322]	<0.0005	1004 [303]	<0.0005	937 [311]	<0.0005	914 [291]	<0.0005
Creatine (3.04 s δ [CH ₃])	46[3–448] [35]	44 [20]	57 [27]	0.015	37 [14]	0.079	35 [9]	0.011	33 [7]	0.003

doi:10.1371/journal.pone.0167309.t001

Discussion

Using the combination of important metabolites identified employing the three statistical methods (Fig 7), the quantified information (Table 1) and biochemical interpretation of the metabolomics data, we constructed a metabolite profile reflecting the primary GLYAT-catalyzed biotransformation of benzoic acid and the metabolic consequences of benzoic acid ingested via flavored water (Fig 8).

Humans conjugate a variety of aliphatic and aromatic monocarboxylic acids with several amino acids; the resulting peptides are excreted in the urine and bile. In characterizing two closely related GLYATs, Nandi *et al.* [18] indicated through kinetic studies that benzoyl-CoA is the main substrate for a benzoyltransferase, with salicyl-CoA and certain aliphatic acyl-CoAs being lesser substrates for it as well. The closely related phenylacetyltransferase utilize phenylacetyl- and indoleacetyl-CoA as substrates. Acyl-CoA substrates of one transferase did not serve as substrate for the other, but act as competitive inhibitors. Glycine is the preferred acyl acceptor for both enzymes, with K_m^{app} for benzoyltransferase being less (3 mM glycine) than for the phenylacetyltransferase (20 mM glycine). Due to the high activity of glutamine N-phenylacetyltransferase, phenylacetylglutamine is almost never detected in human urine. Almost all the phenylacetic acid is excreted as N-phenylacetylglutamine. It thus may be anticipated that formation of hippuric acid in this intervention study mainly reflects the GLYAT activity of the benzoyl variant (EC 2.3.1.71).

Early studies showed that rat liver mitochondria synthesize hippuric acid at a rate of up to 4 nmol/min per mg of protein [29] and comparative kinetic analysis suggested that the formation of the benzoyl-CoA substrate is the rate-limiting factor. More recently, two distinct forms of xenobiotic/medium-chain fatty acid:CoA ligase (XM-ligase) were isolated from human liver mitochondria, referred to as HXM-A and HXM-B [30]. Both forms had medium-chain fatty acid:CoA ligase activity but HXM-A showed 60–80% activity towards 15 different carboxylic acids relative to benzoic acid, its best xenobiotic substrate (100% activity and the highest V_{max}/K_m). Hexanoic acid was the best substrate for HXM-B, although it was also active towards xenobiotic carboxylic acids. In accordance with these findings we propose that medium-chain acyl-CoA ligase (EC 6.2.1.2) is a major catalyst for production of the CoA-substrate required for hippuric acid formation in the present study.

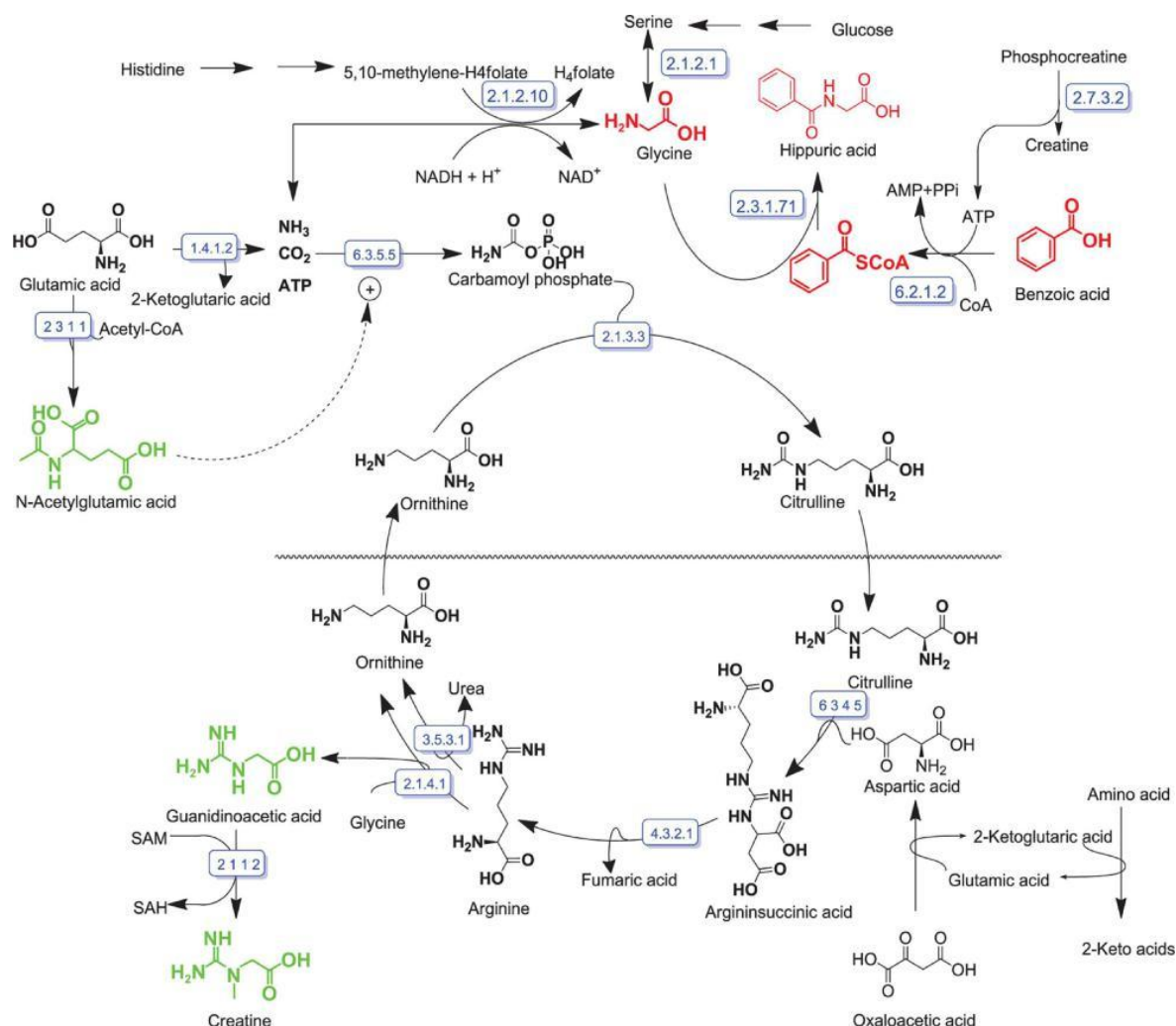


Fig 8. Metabolite profile of benzoic acid biotransformation with hippuric acid as outcome. The metabolites of the primary biotransformation reactions are shown in red. Metabolites of metabolic reactions proposed as being associated with the biotransformation reactions are shown in black. [Metabolites shown in green are presumed to decrease due to increased glycine demand and synthesis of glycine through the reversed glycine cleavage system.] Enzyme nomenclature (names accepted by the IUBMB): EC 1.4.1.2: glutamate dehydrogenase; EC 2.1.1.2: guanidinoacetate N-methyltransferase; EC 2.1.2.1: glycine hydroxymethyltransferase (alt.: serine hydroxymethyltransferase); EC 2.1.2.10: aminomethyltransferase (a glycine synthetase); EC 2.1.3.3: ornithine carbamoyltransferase; EC 2.1.4.1: glycine amidinotransferase; EC 2.3.1.1: amino-acid N-acetyltransferase; EC 2.3.1.71: glycine N-benzoyltransferase (a GLYAT); EC 2.7.3.2: creatine kinase; EC 3.5.3.1: arginase; EC 4.3.2.1: argininosuccinate lyase; EC 6.2.1.2: butyrate-CoA ligase (a medium-chain acyl-CoA ligase); EC 6.3.4.5: argininosuccinate synthase; EC 6.3.5.5: carbamoyl-phosphate synthase (glutamine-hydrolyzing)

doi:10.1371/journal.pone.0167309.g008

Acyl-CoA ligases/synthetases belong to a superfamily of adenylate-forming enzymes, and catalyze the two-step activation of fatty acids or carboxylate-containing xenobiotics [31], with xenobiotic-CoA formation in parallel to endogenous fatty acid activation through the role of the ATP-dependent acid:CoA ligases [32]. We visualize the activation of benzoyl-CoA through a Bi Uni Uni Bi Ping-Pong molecular mechanism proposed for CoA-substrate formation by a long-chain fatty acyl-CoA synthetase [33]: The benzoyl carboxylate substrate (BA) first reacts with enzyme-bound ATP to form an acyl-adenylate intermediate (Ec:B~AMP), which then reacts with CoA to produce the activated benzoyl-CoA ester (B~CoA), having pyrophosphate (PPi) and AMP as byproducts.

Formation of hippuric acid, following GLYAT-catalyzed conjugation between benzoyl-CoA and glycine, peaks within one hour ($p < 0.0001$) following the consumption of flavored water (Table 1). The p -values of the Wilcoxon signed rank test (Table 1) for the time-dependent urinary excretion of glycine complements that of hippuric acid, and likewise peaks at one hour ($p = 0.001$) following the intervention. However, the excretion kinetics profile of glycine clearly suggests abundance of glycine during the main detoxification period (two hours after the intervention), rather than glycine depletion. Two main pathways exist for increased glycine biosynthesis (Fig 8) under physiological conditions of glycine demand: (1) L-serine can be converted to glycine by serine hydroxymethyltransferase (EC 2.1.2.1) in the reversible glycine biosynthesis pathway, having tetrahydrofolate as acceptor for the CH_2OH group from serine, yielding 5,10-methylene-tetrahydrofolate and water. (2) The *de novo* synthesis of glycine can occur from CO_2 and NH_3 , catalyzed by glycine synthase (EC 2.1.2.10) using 5,10-methylene-tetrahydrofolate as the source of the second carbon and yielding tetrahydrofolate — likewise in a reversible reaction. The CO_2 and NH_3 substrates for the *de novo* synthesis of glycine are generated by mitochondrial glutamate dehydrogenase (EC 1.4.1.2) that converts glutamic acid to 2-ketoglutaric acid, with NH_3 usually as a substrate in the urea cycle. Glycine is also a substrate for glycine transaminase (EC 2.1.4.1)-catalyzed synthesis of guanidinoacetic acid. Urinary guanidinoacetic acid concentrations, notably, decreased (Table 1; $p = 0.173$) within the first hour following consumption of flavored water, towards values below the detection limit (Table 1; $p = 0.028$) two hours later.

We speculate that decreased guanidinoacetic acid is caused by the preferential utilization of glycine for benzoic acid detoxification as well as lower urea cycle activity due to decreased N-acetylglutamic acid, a modulator for the up-regulation of the urea cycle. Against this background it seems that the traditional paradigm of GLYAT-catalyzed benzoic acid detoxification, supported by increased *de novo* glycine biosynthesis (Fig 8), prevails under the conditions of the present intervention experiment, even though up-regulation of glycine-amidinotransferase (EC 2.1.4.1) through decreased creatine [34] cannot be excluded.

Finally, we observed an increase in urinary creatine within the first hour following the intervention. The creatine/phosphocreatine system plays an important role in energy storage and energy provision, with creatine synthesis being central in cellular energy metabolism. Two main enzymes are the basis of the creatine biosynthesis pathway, namely, arginine:glycine amidinotransferase (EC 2.1.4.1) and S-adenosyl-L-methionine:N-guanidinoacetate methyltransferase (EC 2.1.1.2), as shown in Fig 8. Given the decrease of guanidinoacetic acid to below the detection limit (Table 1), it seems unlikely that the urinary creatine originates from, and excess creatine is produced by, the creatine biosynthesis pathway. Our proposal is that phosphocreatine (EC 2.7.3.2) degrades in favor of creatine and supplements the ATP reserves, required for the burst of ATP required following benzoic acid detoxification.

In summary, we have described the metabolite profile following benzoic acid intake as part of a designed intervention study. The time-dependent glycine profile supported the view of abundant glycine availability during the main detoxification period rather than that of glycine depletion. It should be noted, however, that the perturbation caused by benzoic acid consumption may be more complex than discussed above. We observed small, but significant time-dependent changes in the NMR spectra (Figure I in S1 File) for methylguanidine and three unknown substances that are not accounted for in the metabolic model shown in Fig 8.

The complex NMR spectral data, generated from cases participating in a time-dependent cross-over study, could be resolved sufficiently through the application of traditional univariate and multivariate analyses combined with an ANOVA-simultaneous component analysis (ASCA), repeated measures analysis of variance (RM ANOVA) and unfolded principal component analysis (unfolded PCA) — an approach that opens a novel way for analyses and understanding of complex metabolomics data that reflect perturbations from normal or pathophysiological endogenous or exogenous origin.

Furthermore the combination of the complementary statistical methods together with the transdisciplinary approach followed in metabolomics research provided a more holistic view of the data. This proved useful in elucidating the effect of the intervention and provided novel insights information into benzoic acid biotransformation, which is not typically observed by more reductionist methods of analysis.

Supporting Information

S1 File.

(DOCX)

S2 File.

(XLSX)

S3 File.

(M)

Acknowledgments

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. CI received a post-graduate bursary from the National Research Foundation (NRF) 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 NRF or TIA funders. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceptualization: CJR JAW.

Data curation: CI SM MvR JAW.

Formal analysis: CI SM.

Funding acquisition: CJR.

Investigation: CI SM.

Methodology: CJR JAW.

Project administration: CJR JAW.

Software: JAW MvR.

Supervision: CJR JAW.

Writing & original draft: CI MvR SM LJM JAW CJR.

Writing & review & editing: CI MvR SM LJM JAW CJR.

References

- Gibney MJ, Walsh M, Brennan L, Roche HM, German B, Van Ommen B. Metabolomics in human nutrition: opportunities and challenges. *Am J Clin Nutr*. 2005; 82:497–503. PMID: [16155259](#)
- Wishart DS. Metabolomics: applications to food science and nutrition research. *Trends Food Sci Tech*. 2008; 19:482–493.
- Antti H, Bollard ME, Ebbels T, Keun H, Lindon JC, Nicholson JK, et al. Batch statistical processing of ¹NMR-derived urinary spectral data. *J Chemometrics*. 2002; 16:461–468.
- Smilde AK, Jansen JJ, Hoefsloot HCJ, Lamers R-JAN., Van der Greef J, Timmerman ME. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics*. 2005; 21:3043–3048. doi: [10.1093/bioinformatics/bti476](#) PMID: [15890747](#)
- Park H-W, Park EH, Yun H-M, Rhim H. Sodium benzoate mediated cytotoxicity in mammalian cells. *J Food Biochem*. 2011; 35:1034–1046.
- Schachter D, Taggart JV. Benzoyl-coenzyme A and hippurate synthesis. *J Biol Chem*. 1953; 203:925–934. PMID: [13084662](#)
- Conti A, Bickel MH. History of drug metabolism: discoveries of the major pathways in the 19th century. *Drug Metab Rev*. 1977; 6:1–50.
- Beyoğlu D, Smith RL, Idle JR. Dog bites man or man bites dog? The enigma of the amino acid conjugations. *Biochem Pharmacol*. 2011; 83:1331–1339. doi: [10.1016/j.bcp.2011.12.031](#) PMID: [22227274](#)
- Beyoğlu D, Idle JR. The glycine deportation system and its pharmacological consequences. *Pharmacol Therapeut*. 2012; 135:151–167.
- Wen H, Yang H-J, An YJ, Kim JM, Lee DH, Jin X, Park S-W, Min K-J, Park S. Enhanced phase II detoxification contributes to beneficial effects of dietary restriction as revealed by multi-platform metabolomics studies. *Mol Cell Proteomics*. 2013; 12:575–586. doi: [10.1074/mcp.M112.021352](#) PMID: [23230277](#)
- Nicholson JK, Holmes E, Wilson ID. Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol*. 2005; 3:431–438. doi: [10.1038/nrmicro1152](#) PMID: [15821725](#)
- Jana S, Mandlekar S. Role of phase II drug metabolizing enzymes in cancer chemoprevention. *Curr Drug Metab*. 2009; 10:595–616. PMID: [19702535](#)
- Zimniak P. Detoxification reactions: relevance to aging. *Ageing Res Rev*. 2008; 7:281–300. doi: [10.1016/j.arr.2008.04.001](#) PMID: [18547875](#)
- Holtzclaw WD, Dinkova-Kostova AT, Talalay P. Protection against electrophile and oxidative stress by induction of phase 2 genes: the quest for the elusive sensor that responds to inducers. *Adv Enzyme Regul*. 2004; 44:335–367. doi: [10.1016/j.advenzreg.2003.11.013](#) PMID: [15581500](#)
- Aprison MH, Werman R. The distribution of glycine in cat spinal cord and roots. *Life Sci*. 1965; 4:2075–2083. PMID: [5866625](#)
- Tanaka I, Ezure K. Overall distribution of GLYT2 mRNA-containing versus GAD67 mRNA-containing neurons and colocalization of both mRNAs in midbrain, pons, and cerebellum in rats. *Neurosci Res*. 2004; 49:1874–1882.
- Badenhorst CPS, Van der Sluis R, Erasmus E, Van Dijk AA. Glycine conjugation: importance in metabolism, the role of glycine N-acyltransferase, and factors that influence interindividual variation. *Expert Opin Drug Met*. 2013; 9:1139–1153.
- Nandi DL, Lucas SV, Webster LT Jr. Benzoyl-Coenzyme A:Glycine N-acyl-transferase and Phenylacetyl-Coenzyme A:Glycine N-Acyltransferase from Bovine Liver Mitochondria. *J Biol Chem*. 1979; 254:7230–7237. PMID: [457678](#)
- Schachter D, Taggart JV. Glycine N-acylase: purification and properties. *J Biol Chem*. 1954; 208:263–275. PMID: [13174534](#)
- Lindon JC, Holmes E, Nicholson JK. Pattern recognition methods and applications in biomedical magnetic resonance. *Prog Nucl Magn Reson Spectrosc*. 2001; 39:1–40.
- Villez K, Steppe K, De Pauw DJW. Use of unfold PCA for online plant stress monitoring and sensor failure detection. *Biosystems Engineering*. 2009; 103:23–24.
- Mason S. The metabolomics of acute alcohol abuse. M.Sc. Thesis, North-West University. 2010. Available: <http://dspace.nwu.ac.za/handle/10394/5034>
- Fave  G, Beckmann M, Lloyd AJ, Zhou S, Harold G, Lin W, et al. Development and validation of a standardized protocol to monitor human dietary exposure by metabolite fingerprinting of urine samples. *Metabolomics*. 2011; 7:469–484. doi: [10.1007/s11306-011-0289-0](#) PMID: [22039364](#)
- Viant M, Bearden DW, Bundy JG, Burton IW, Colette TW, Ekman DR, et al. International NMR-based environmental metabolomics intercomparison exercise. *Environ Sci Technol*. 2009; 43:219–225. PMID: [19209610](#)
- Dona AC, Jime nez B, Schaa fer H, Humpfer E, Spraul M, Lewis MR, et al. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal Chem*. 2014; 86(19):9887–9894. doi: [10.1021/ac5025039](#) PMID: [25180432](#)

26. Ellinger JJ, Chylla RA, Ulrich EL, Markley JL. Databases and Software for NMR-based metabolomics. *Curr Metabolomics*. 2013; 1:28–40.
27. Powers R. NMR Metabolomics and drug discovery. *Magn Reson Chem*. 2009; 47, S2–S11. doi: [10.1002/mrc.2461](https://doi.org/10.1002/mrc.2461) PMID: [19504464](https://pubmed.ncbi.nlm.nih.gov/19504464/)
28. Zwanenburg G, Hoefsloot HCJ, Westerhuis JA, Jansen JJ, Smilde AK. ANOVA–principal component analysis and ANOVA–simultaneous component analysis: a comparison. *J Chemometrics*. 2011; 25:561–567.
29. Gatley SJ, Sherratt HAS. The synthesis of hippurate from benzoate and glycine by rat liver mitochondria. *Biochem J*. 1977; 166:39–47. PMID: [901416](https://pubmed.ncbi.nlm.nih.gov/901416/)
30. Vessey DA, Kelley M, Warren RS. Characterization of the CoA ligases of human liver mitochondria catalyzing the activation of short and medium-chain fatty acids and xenobiotic carboxylic acids. *Biochim Biophys Acta*. 1999; 1428:455–462. PMID: [10434065](https://pubmed.ncbi.nlm.nih.gov/10434065/)
31. Kochan G, Pilka ES, von Delft F, Oppermann U, Yue WW. Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A. *J Mol Biol*. 2009; 388:997–1008. doi: [10.1016/j.jmb.2009.03.064](https://doi.org/10.1016/j.jmb.2009.03.064) PMID: [19345228](https://pubmed.ncbi.nlm.nih.gov/19345228/)
32. Knights KM, Drogemuller CJ. Xenobiotic-CoA Ligases: Kinetic and Molecular Characterization. *Curr Drug Metab*. 2000; 1:49–66. PMID: [11467080](https://pubmed.ncbi.nlm.nih.gov/11467080/)
33. Hisanaga Y, Ago H, Nakagawa N, Hamada K, Ida K, Yamamoto M, et al. Structural basis of the substrate-specific two-step catalysis of long chain fatty acyl-CoA synthetase dimer. *J Biol Chem*. 2004; 279:31717–31726. doi: [10.1074/jbc.M400100200](https://doi.org/10.1074/jbc.M400100200) PMID: [15145952](https://pubmed.ncbi.nlm.nih.gov/15145952/)
34. McGuire DM, Gross MD, Van Pilsum JF, Towle HC. Repression of rat kidney L-arginine:glycine amidinotransferase synthesis by creatine at a pretranslational level *J Biol Chem*. 1984; 259:12034–12038.
35. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P, Dame ZT, Poelzer J, Huynh J, Yallou FS, Psychogios N, Dong E, Bogumil R, Roehring C, Wishart DS. The human urine metabolome. *PLOS ONE*. 2013; 8(9):e73076. doi: [10.1371/journal.pone.0073076](https://doi.org/10.1371/journal.pone.0073076) PMID: [24023812](https://pubmed.ncbi.nlm.nih.gov/24023812/)
36. Fingerhut R. Stable isotope dilution method for the determination of guanidinoacetic acid by gas chromatography/mass spectrometry. *Rapid Commun Mass Spectrom*. 2003; 17(7):717–722 doi: [10.1002/rcm.966](https://doi.org/10.1002/rcm.966) PMID: [12661026](https://pubmed.ncbi.nlm.nih.gov/12661026/)

Supporting Information

Contribution towards a Metabolite Profile of the Detoxification of Benzoic Acid through Glycine Conjugation: An Intervention Study

Cindy Irwin^{1,¶}, Mari van Reenen^{1,2,¶}, Shayne Mason¹, Lodewyk J. Mienie¹, Johan A. Westerhuis^{2,3} and Carolus J. Reinecke^{1,*}

¹Centre for Human Metabolomics, Faculty of Natural Sciences, North-West University (Potchefstroom Campus), Potchefstroom, South Africa

²Department of Statistics, Faculty of Natural Sciences, North-West University (Potchefstroom Campus), Potchefstroom, South Africa

³Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands

* Corresponding author

E-mail: carools.reinecke@nwu.ac.za

¶ These authors contributed equally to this work

Email addresses:

Cindy Irwin:	cindyirwin74@gmail.com
Mari van Reenen:	12791733@nwu.ac.za
Shayne Mason:	nmr.nwu@gmail.com
Lodewyk J. Mienie:	japie.mienie@gmail.com
Johan A. Westerhuis:	J.A.Westerhuis@uva.nl
Carolus J. Reinecke:	carools.reinecke@nwu.ac.za

Table of Contents

Section 1: Original ^1H -NMR spectral data for intervention 1	3
Section 2: Threshold value and normalization	4
Section 3: Data pre-processing and statistical analysis	4
Section 3.1: Case reduction	5
Section 3.2: Data transformation and scaling	7
Section 3.3: Cross-sectional analysis	7
Section 3.3.1: Univariate analysis	7
Section 3.3.2: Multivariate analysis	9
Section 3.4: RM ANOVA	14
Section 3.5: Unfolded PCA	15
Section 3.6: ASCA	16
Section 3.7: Graphs on excretion kinetics	18
Section 3.8: NMR spectra on the excretion of six substances	19
Section 4: NMR analysis on guanidinoacetic acid	20
References	22

Section 1 Original ¹H-NMR spectral data for intervention 1

Intervention 1 refers to consumption of flavored water only – the cause of the benzoic acid biotransformation described in the main paper. The data include urine samples from the 23 experimental cases that completed the study, yielding a total of 138 study samples. The original design was based on 24 cases (individual subjects), although one did not participate in the flavored water experiment. The original ¹H-NMR spectrum of each of the urine samples analyzed was divided into 0.02-ppm equal-sized bins between 0.5 and 10 ppm, excluding the region of the water peak (4.72–4.88 ppm). This resulted in 467 bins containing spectral data. The original, raw spectral data matrix of all the samples from the vehicle intervention analyzed for this study are given as an electronic file in Excel format (S2 File: CI_unscaled raw vehicle data.xlsx). Table A represents a small extract from this file.

					Bin	1	2	3	...	466	467
					No.	9.99	9.97	9.95	...	0.53	0.51
Batch	Sample name	Case	Treatment	Time							
1	F41-1	F	Vehicle	–1		252890.2	171823.4	163215.5		384216.4	266010.8
1	F410	F	Vehicle	0		151299.1	100954.3	86908.15		262467	216768.2
1	F411	F	Vehicle	1		79268.87	105579.2	95048.77		76325.53	126937.3
1	F412	F	Vehicle	2		84927.78	101751.3	127028.2		77872.51	68765.72
1	F413	F	Vehicle	3		74747.86	144056.7	47937.61		91158.41	58079.11
1	F414	F	Vehicle	4		67488.71	70024.98	68237.62		103052	89226.37
2	E51-1	E	Vehicle	–1		120268.5	125437.8	123530.6		143314	122366.1
2	E510	E	Vehicle	0		206635.9	142868.1	112930.5		198141.4	156402.8
2	E511	E	Vehicle	1		106837.9	97280.69	132238.9		168890.4	128759.2
2	E512	E	Vehicle	2		88524.67	110966.2	92268.4		75853.71	102751.3
2	E513	E	Vehicle	3		101023.8	97107.34	59599.65		76011.03	88203.6
2	E514	E	Vehicle	4		134874.8	93287.56	67920.9		182858.7	85690.59
3	G41-1	G	Vehicle	–1		183044.6	110147.5	92701.17		368924.5	308433.9
3	G410	G	Vehicle	0		220505.4	134572.8	99752.4		451830.2	324699.9
3	G411	G	Vehicle	1		229915.6	116229.6	130475.5		951134.8	407340.4

Table A: Raw ¹H-NMR spectral data are given as an electronic file in Excel format (see S2 File: CI_unscaled raw vehicle data.xlsx) attached online as part of the Supporting Information

Section 2 Threshold value and normalization

The use of NMR spectroscopy as a tool for metabolomics is limited by the sensitivity (limit of detection) of NMR [1], requiring an approach in discarding noise while retaining and then measuring real peaks. The spectral intensities within the original 467 bins generated here ranged from 7.8×10^5 to 1.7×10^9 . Based on previous empirical experience with NMR spectral analysis, we defined a threshold value of 2×10^6 , being approximately the limit of detection of metabolomic substances presumed to be present in a spectral bin. To reduce the uncertainty in the data at the level of the approximate detection limit, we applied the threshold of 2×10^6 to the data matrix, setting all values below this threshold to zero. We subsequently applied a “zero-filter” to remove bins with more than 50% zero observations

across all six time points, as applied for fusion of MS-based metabolomics data [2]. This resulted in a reduction in the number of bins containing spectral data from 467 to 347. To account for the dilution differences between the urine samples, the value of each bin was made relative to the CH₂ peak of creatinine. This peak is contained in the 4.05 and 4.07 ppm bins. Therefore, each bin value was divided by the sum of the value in these two bins to normalize the bin values relative to creatinine. After normalization, the three bins related to creatinine (4.05, 4.07 and 3.05 ppm) were also removed from the matrix, giving a final total of 344 bins containing spectral data. The above processing was performed in Matlab [3] as outlined in the script provided as additional material.

Section 3 Data pre-processing and statistical analysis

The data resulting from this intervention study potentially contain volumes of information. However, unlocking it is a challenge for various reasons, the first being the presence of three major sources of variation: (i) the four experimental treatments; (ii) the variation over time; and (iii) the variation between cases participating in the study (potentially the greatest source of variation). The second relates to the analytical platform used. NMR spectra using equidistant binning results in hundreds of bins, in this instance 467. This far exceeds the number of cases measured (23), which is known to cause more traditional statistical methods such as MANOVA (multivariate analysis of variance) to fail [4]. Third, the quantified values may be very low for some bins and carry large uncertainty as to the accuracy of the value reported. As a result, values below a certain threshold are converted to zeros (Section 2 above), which again poses a challenge for many statistical methods making an assumption about distribution.

The design of the study aimed to alleviate these challenges to some extent by (i) repeatedly measuring the same individuals across both time and treatment factors; (ii) selecting cases with similar demographics, i.e. age and living conditions; and (iii) limiting the influence of dietary difference by exposing cases to each intervention while in a fasting state. Finally, the results discussed in this paper pertain only to a subset of the data, one of the four treatment interventions, making this initial analysis more palatable yet still highly informative.

Taking into account the design of the experiment as well as the limitations of the various statistical techniques, a range of methods were applied to this subset of the data, as discussed in more detail below.

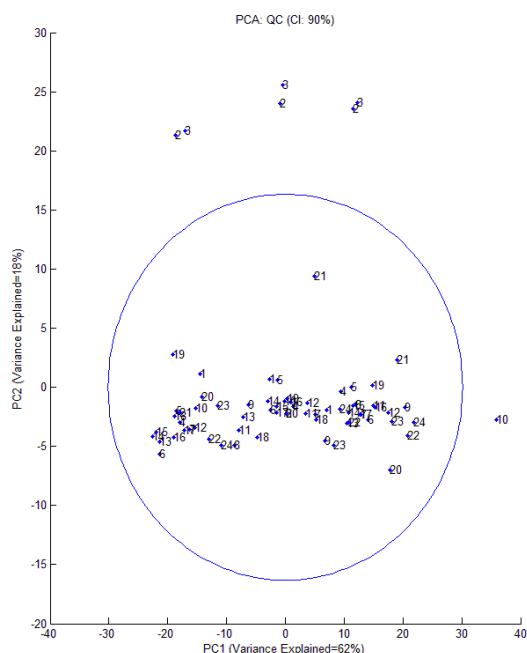
Section 3.1 Case reduction

Reproducibility and repeatability are not known concerns in NMR analysis [5]. Nevertheless, three pooled quality control (QC) samples were included in each NMR batch run, in the first, middle and last positions. The role of the QC samples was not to correct for any drift or between-batch effects, as done in GC–MS for instance, as this category of confounding factors should not be present in NMR analysis. Rather, the QC samples were used to identify batches which produced unexpected and therefore untrustworthy results due to some external factor we did not or could not control or measure. Outlier detection was based on the scores from a principal component analysis (PCA) with 90% confidence regions, along with Hotelling's T^2 statistic (Figure A). A two-component PCA model was

constructed based on the log-scaled and centered normalized data (464 bins) from only the QC samples. Note that the 50% zero filter was not yet applied as some of the zero measurements may be due to outlying batches.

All three QC samples from batches 2 and 3 were identified as outliers. This could have been due to some analytical error which occurred during the preparation of the samples from these two batches. This chemical shift in the NMR spectra of the samples in these two batches, however, made it difficult to compare the peaks in these samples with those of the samples from the 21 other batches. Consequently, these two batches were excluded from further analysis as batch corrections based on only three QCs may introduce more error than they remove. We could eliminate these batches with little effect as all samples from a given case represented a batch; so, even though we sacrificed in terms of sample size, the exclusions did not change the design.

A



B

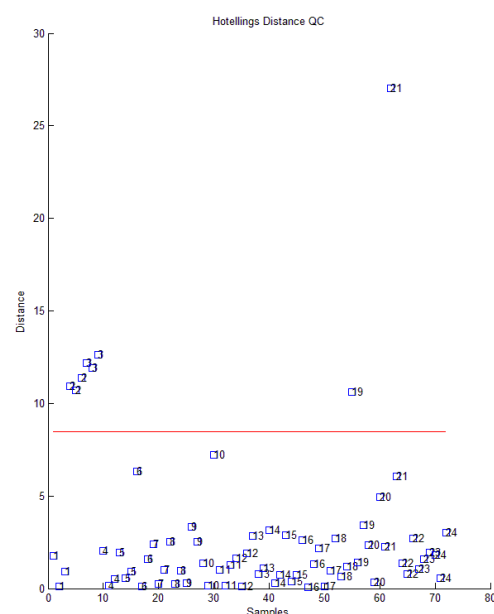


Figure A: QC Outlier Detection

The figure shows the results of a PCA analysis on the QC samples. The PCA scores of all three QC samples for batches 2 and 3 fall outside the 90% confidence region (A). The Hotelling's T^2 statistics for these same samples flag them as outliers (B). Two other QC samples from different batches were also flagged; however, as only one of the three QCs in these batches was flagged, these batches were not removed.

The above processing and analysis was performed in Matlab using PCA function provided by the PLS Toolbox [3], as outlined in the script provided as additional material.

Section 3.2 Data transformation and scaling

To reduce the uncertainty in the data, we applied a threshold to the data matrix followed by a "zero-filter" (Section 2). The reduced data were log transformed to ensure that the additive

models used to extract information from the data were appropriate [6]. Log transformation also makes skew distributions – often resulting when variables (i.e. bins) assume only positive values, such as here – more symmetric. Symmetry is assumed in many statistical methods, including some of the methods used here. The above processing was performed using Matlab.

Section 3.3 Cross-sectional analysis

Here we made use of various methods to compare times –1, 1, 2, 3 and 4 hours with time 0. We did this to answer the following questions:

- (i) At what time did the effect of the vehicle first become apparent?
- (ii) At what time did the effect of the vehicle start to diminish?
- (iii) Were individuals in a fasting state more similar?
- (iv) Did their metabolism change greatly while fasting?

While the last two questions do not relate to the aims of the study, their answers influence how we interpret other results. Next, we discuss the univariate methods used. These analyses were performed in Matlab and included (i) univariate statistics using the statistics toolbox; (ii) PCA using the PLS Toolbox; and (iii) PLS–DA using the PLS Toolbox [3].

Section 3.3.1 Univariate analysis

The Wilcoxon signed rank test was first employed to find significant differences in the average ranks of bin values between two time points. This hypothesis test is non-parametric (i.e. makes no distributional assumptions) and assumes samples are paired. Since cases were measured repeatedly, it is important to use statistical methods that can model this dependence. Paired measures reduce some of the variation that is not of interest (i.e. between cases) and methods exploiting this are therefore better able to detect group differences. All hypothesis tests rely on random sampling, which is not the case here. Convenience sampling, as used here, can bias p-values produced by tests and they should be evaluated with caution. Further multiple testing requires controlling the family-wise error rate. However, since we did not base variable selection on hypothesis testing, these factors were less of a concern. Our aim here was simply to gain insight into the general progression of the effect of the treatment over time (i.e. whether differences became more pronounced or not).

Given the limitations of hypothesis testing listed above, other univariate measures supplemented our evaluation of the general progression. For each bin, two more intuitive measures of group differences were also evaluated: (i) effects sizes based on the Wilcoxon signed rank z-statistic to evaluate practical significance; and (ii) fold change value.

Figure B provides the Volcano plots for all pairwise time point comparisons. A Volcano plot is essentially a scatter plot of the $\log_2$ scaled fold change values against the $-\log_{10}$ scaled p-values (in this case p-values from the Wilcoxon signed rank test). By performing scaling, the plot is able to highlight the discriminatory bins in the upper left and right corners for a given threshold. In this instance bins with p-values below 0.05 and fold change values larger than 2 were highlighted in the figures. From Figure B it is evident that individual bins start to reflect the effect of the vehicle intervention 1 hour after consumption.

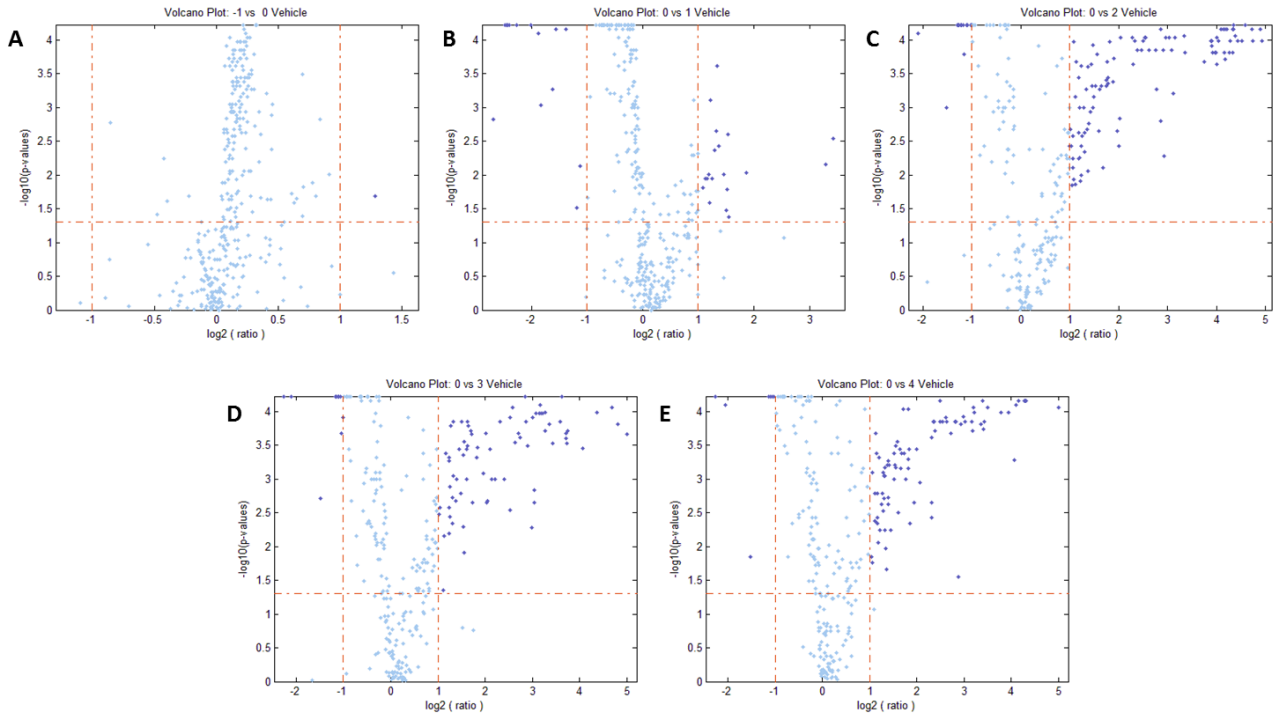


Figure B: Volcano Plots of Pairwise Time Point Comparisons with time 0

The figure displays the individual Volcano plots for the pairwise comparisons of time -1 vs. time 0 (A); time 0 vs. time 1 (B); time 0 vs. time 2 (C); time 0 vs. time 3 (D); and time 0 vs. time 4 (E).

Section 3.3.2 Multivariate analysis

To gain insight into the interactions between bins, i.e. to find bins which in combination differ between groups, we made use of multivariate statistical methods. Euclidean and Ward cluster analysis was used to find natural groupings between the cases based on the bin data provided, but without providing information on the known group structure (unsupervised). Cluster analysis makes use of a distance measure (Euclidean) to measure the distances or difference between samples based on the measured variables. A linkage method (Ward) was then used to group neighboring samples. The groups formed can be viewed using a dendrogram. If these groups coincide with the known group structure (as indicated by labeling the dendrogram after the fact), we can be confident that the data set contains discriminatory bins. For more detail on cluster analysis refer to Field [7]. Figure C shows the dendrograms for all pairwise time point comparisons with time 0. The interesting findings from these dendrograms are discussed in the main paper.

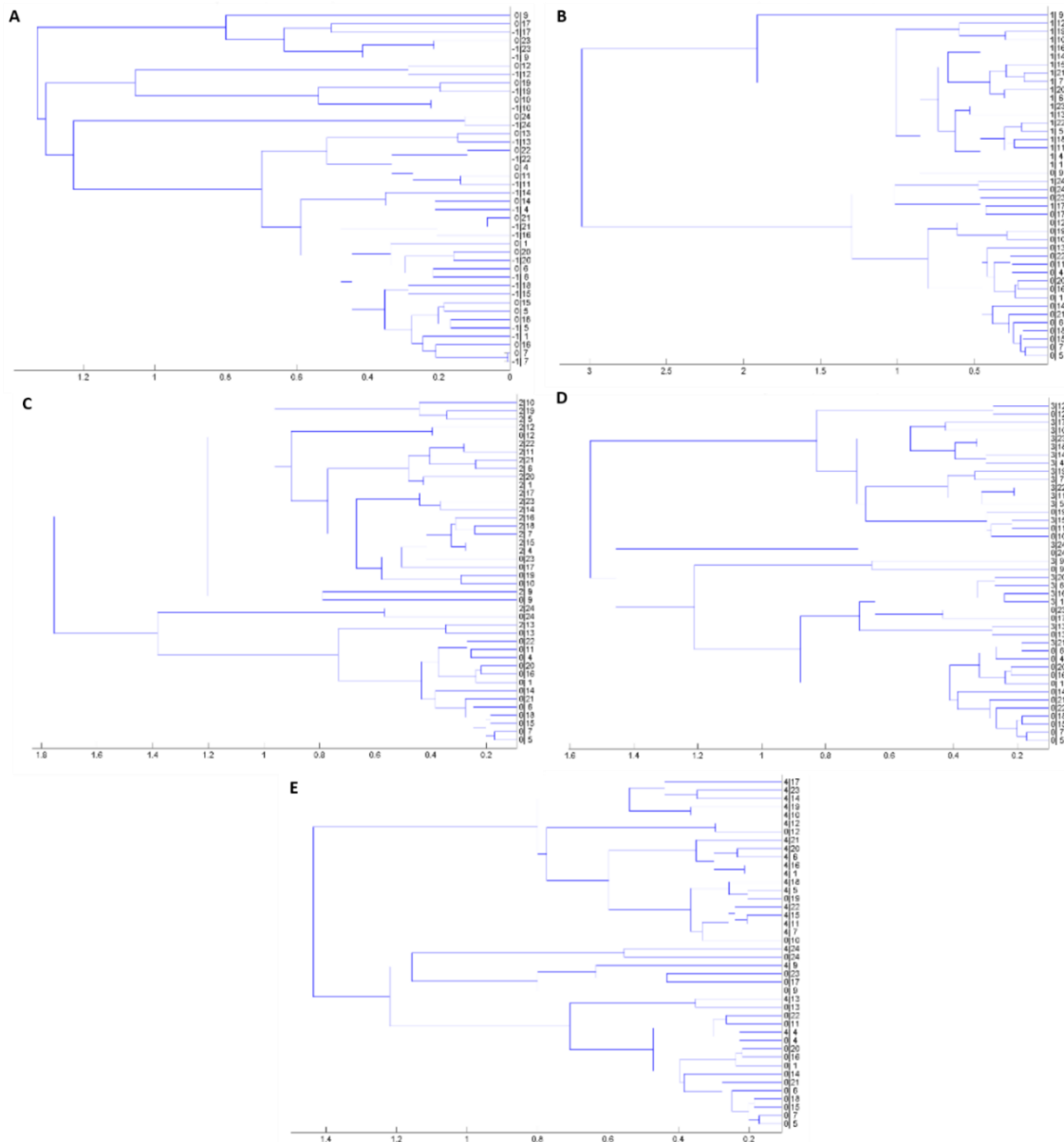


Figure C: Dendrograms of Pairwise Time Point Comparisons with time 0

The figure displays the individual dendrograms for the pairwise comparisons of time -1 vs. time 0 (A); time 0 vs. time 1 (B); time 0 vs. time 2 (C); time 0 vs. time 3 (D); and time 0 vs. time 4 (E).

Principal component analysis (PCA) was used to project the hundreds of bins onto a two-dimensional subspace using the following model:

If D represents a data matrix of N cases (i.e. samples) by K variables, where the N cases represent observations across all factor levels, then the PCA model of D can be written as: $D = TP' + E$, where T represents the scores and P' the transposed loadings matrices, respectively. Given that the dimensionality of D is reduced, the error matrix E is added to account for the variation not explained by the model (TP'). For more detail refer to [8].

Such a projection allows for a clearer view of the influence of experimental factors, specifically which factor dominates the variability in the data. PCA, like cluster analysis, is also an unsupervised method, therefore the influence of experimental factors becomes

visible when labeling the scores (values for samples in the new subspace) according to associated factor level after the model has been constructed, in our case the times associated with each sample. Projecting the matrix also allows us to evaluate the influence of each variable on the projected scores. In other words, if we find that our group structure predominantly explains the covariance structure of the data, we can evaluate the loadings (values for variables or bins in the new subspace) to identify a list of variables with discriminatory ability on their own or in combination. The loadings were not used during the PCA analysis but extensively so in the ASCA model (discussed later), which is built on the PCA model.

Figure D shows the score plots for all pairwise time point comparisons with time 0. Figure D(A) indicates how fasting state did not change over time since times –1 and 0 are very similar and the individual cases are closely comparable. Figures D(B) – D(E) show how the global bin profile changes relative to the starting point of the intervention. The effect of the vehicle is already evident after the first hour, starts to diminish after two hours, and cases continue to return to normal three and four hours after the intervention.

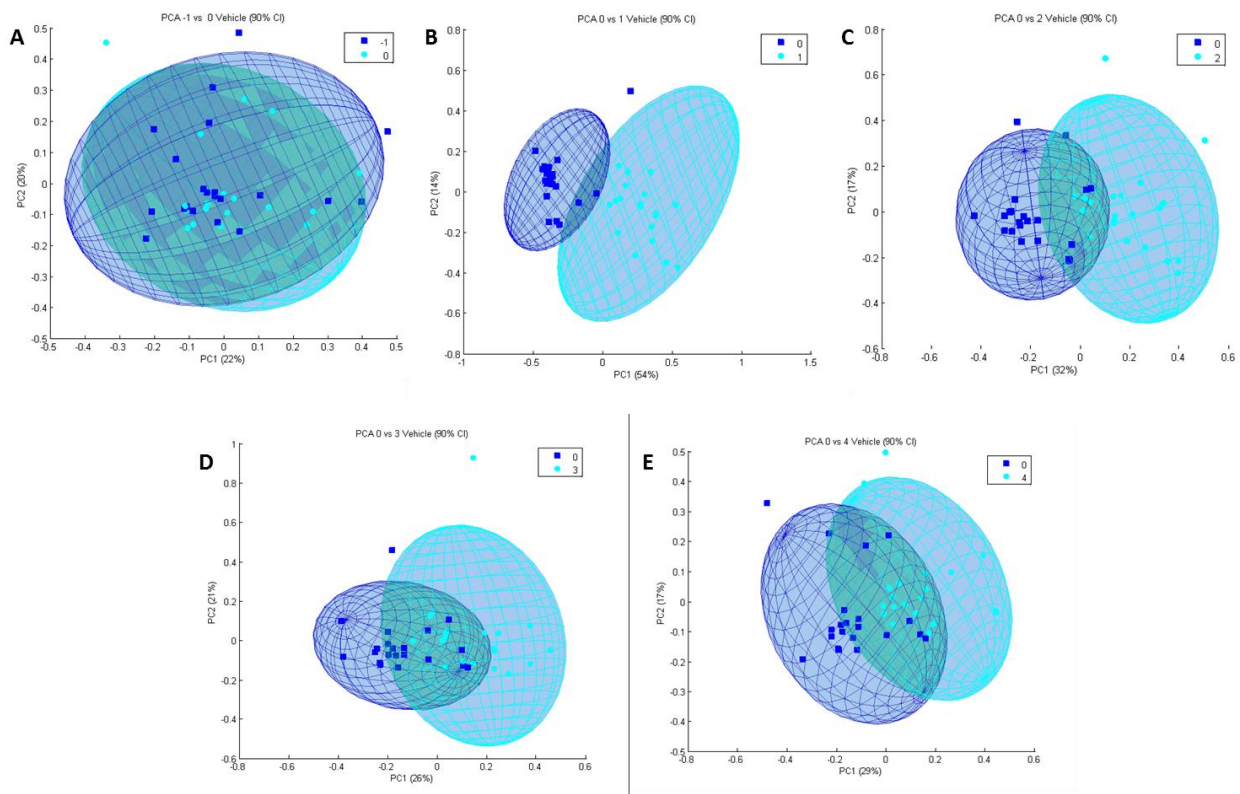


Figure D: PCA Score Plots of Pairwise Time Point Comparisons with time 0

The figure displays the individual score plots of the PCA models for the pairwise comparisons of time –1 vs. time 0 (A); time 0 vs. time 1 (B); time 0 vs. time 2 (C); time 0 vs. time 3 (D); and time 0 vs. time 4 (E).

Lastly, partial least-squares discriminant analysis (PLS–DA) was performed. This method is similar to PCA but makes use of the known group structure (i.e. it is a supervised method). The benefit of using this information is that the method is better able to identify a list of discriminatory variables. The downside is that the model itself may not generalize to other data sets as it is known to overfit [9]. Since our aim was not to develop a classification

model, we were not too concerned about this. However, we do report two measures of confidence in the bins selected: (i) R^2 to measure the goodness of the model fit; and (ii) Q^2 , the leave-one-out cross-validated R^2 , to assess generalizability or sensitivity to the individual samples used to construct the model. Figure E shows the score plots for all pairwise time point comparisons with time 0, whereas Table B contains the fit statistics. We can draw the same conclusion as from the PCA, but with a more pronounced difference, which is to be expected given that PLS–DA is a supervised method.

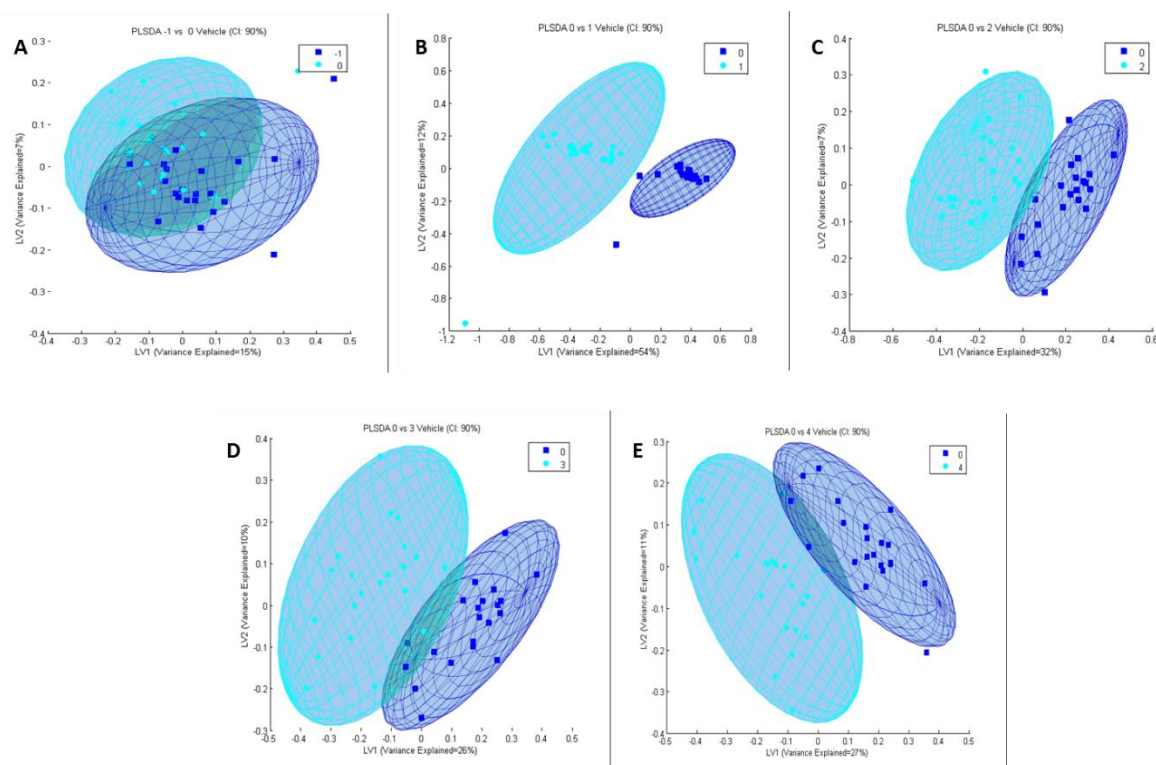


Figure E: PLS–DA Score Plots of Pairwise Time Point Comparisons with time 0

The figure displays the individual score plots of the PLS–DA models for the pairwise comparisons of time –1 vs. time 0 (A); time 0 vs. time 1 (B); time 0 vs. time 2 (C); time 0 vs. time 3 (D); and time 0 vs. time 4 (E).

Comparison	R^2	Q^2
Time 0 vs. –1	0.171	0.002
Time 0 vs. 1	0.776	0.836
Time 0 vs. 2	0.726	0.704
Time 0 vs. 3	0.646	0.488
Time 0 vs. 4	0.628	0.501

Table B: PLS–DA Fit Statistics Pairwise Time Point Comparisons with time 0

The table reports the fit statistics for the first latent variable of PLS–DA models for the pairwise comparisons.

The results from the cross-sectional analysis led us to conclude that, while between-case variation is significant, we were still able to find bins that differ from their baseline measure (time 0). We also found that the fasting state was long enough to reduce between-case variation (time –1 vs. 0). We made use of the bins selected by the PLS–DA models as

one of the shortlists to be included in the Venn analysis that produced the key metabolites quantified and discussed. We wanted to include bins in this list that were informative at any point in time (compared to the baseline, i.e. time 0) while accounting for the multiple PLS–DA models constructed. As such we decided to include bins with VIP values exceeding 2 in any of the models. A VIP score is calculated for each variable in the data matrix and provides an indication of the variable's importance in the projection (hence abbreviated as VIP). For more detail on this calculation refer to [10].

Section 3.4 RM ANOVA

RM ANOVA (repeated measures analysis of variance) is a statistical method that deconstructs a data set into the known sources of variation in order to assess which of these sources significantly influenced the values recorded in the data set. This deconstruction is performed by calculating and subtracting factor level means (ANOVA) – in our case the means for the time points, as well as for each case since the same cases were measured across all time points. The primary aim of this analysis was to identify variables that changed significantly in time (across the six time points). We constructed the ANOVA model as follows:

Let X represent all observations for a given bin and let α and β represent the Time and Case factors, respectively. We can model the change in time as follows:

$$X_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$$

where μ represents the overall mean of the bin

α_i represents the time effects with $i = 1 \dots 6$

β_j represents the case effects with $j = 1 \dots 21$

$\alpha\beta_{ij}$ represent the random interaction effect between case i , treatment j , i.e. the random error not explained by the model.

Therefore X_{ij} represents the observation on the bin for case j , $i - 2$ hours after consuming the vehicle, as modeled on the basis of the factor means and random noise. Bins for which the factor means play a larger role than the random noise to obtain the observed value are flagged as significant.

RM ANOVA essentially expands the cross-sectional approach to account for the complete study design of the data set (or subset since we retain only time as a factor). However, it does not account for interactions between bins as it is a univariate method. RM ANOVA can be expanded to evaluate multiple variables, but breaks down when the number of variables exceeds the number of cases, as mentioned earlier. For in-depth insight into ANOVA methods, refer to Tabachnick & Fidell [11,12]. Here we used RM ANOVA to identify individual bins that changed significantly in time, as constituting the second shortlist to be included in the Venn analysis. To control the family-wise error rate resulting from applying RM ANOVA hypothesis testing multiple times, bins were selected only if their associated p-values were estimated as 0. This analysis was performed in Matlab using the statistics toolbox [3] and code adapted from Trujillo-Ortiz *et al.* [13].

Section 3.5 Unfolded PCA

Unfolded PCA was applied to gain insight into the global (i.e. multivariate) effect of the vehicle over time. This method extends the RM ANOVA idea to the multivariate setting. However, it does not explicitly take the study design into account as it does not truly decompose the data into the known sources of variation. The data tensor (with three dimensions representing cases, times and bins) was unfolded in time as illustrated in Figure F. Unfolding transforms a three-dimensional tensor into a two-dimensional matrix, allowing for principal component analysis. PCA of the unfolded tensor provides insight into the effect of the vehicle in time on the bins (indicated by the ellipses and centroids) as well as individual cases (indicated by the overlaid trajectories) (Figure 5 in the main paper). For a more in-depth discussion on unfolding refer to [14]. This analysis was performed in Matlab using the PCA function of the PLS toolbox.

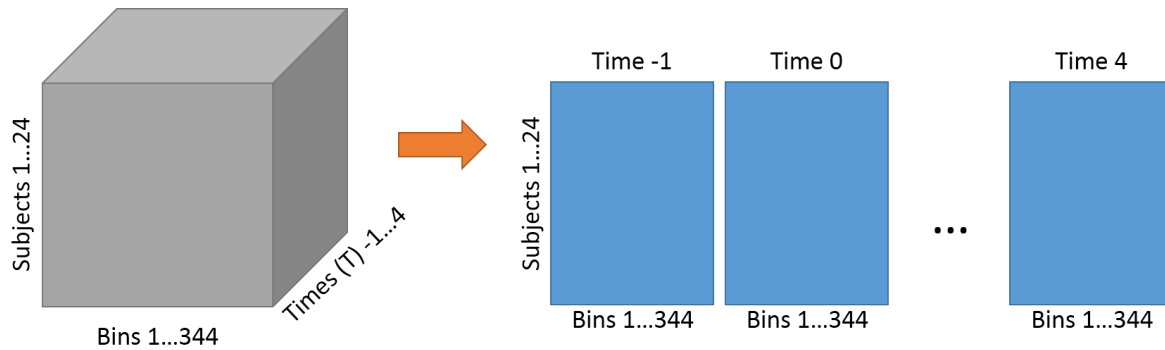


Figure F: Unfolding Data

The figure illustrates how a new data matrix was formed by unfolding the data tensor, generated by the vehicle intervention, over time.

Section 3.6 ASCA

ASCA (ANOVA-simultaneous component analysis) was developed for analyzing designed metabolomics data. It can not only account for the study design but also for the high-dimensionality of the data. In essence ASCA combines MANOVA and PCA (as can be deduced from the name) by first deconstructing the data as done in ANOVA and subsequently applying PCA to each subset. The model can be summarized as follows:

If D represents a data tensor then ASCA will deconstruct D similarly to RM ANOVA:

$$\mathbf{D} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\alpha}\boldsymbol{\beta}_{ij}$$

Note the adoption of boldface symbols, as the factor now represents matrices of means across all variables as opposed to just a vector of means for one variable for each factor level. Next, PCA is performed for each matrix

$$D = 1m + T_{\alpha}P'_{\alpha} + T_{\beta}P'_{\beta} + T_{\alpha\beta}P'_{\alpha\beta}$$

For an in-depth discussion on the ASCA method, refer to Jansen *et al.* [15] and Smilde *et al.* [16]. As mentioned in the PCA discussion, we can evaluate the loadings

(values for variables in the new subspace) to identify a list of bins with discriminatory ability on their own or in combination. The sum of the squared loadings (SSL) for the first two latent variables was used to identify perturbed bins because in combination they explained 88% of the variation. Bins were seen to have discriminatory ability if their SSL values were above 0.01, the threshold identified based on a Scree plot of all SSL scores (Figure G).

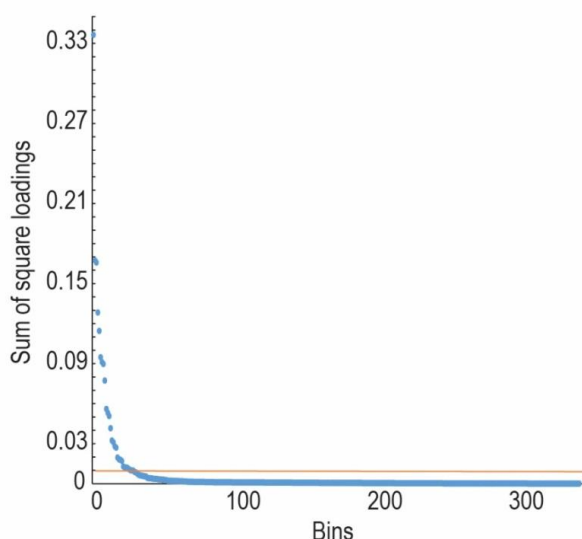


Figure G: Sum of Squared Loadings of ASCA model

A scree plot of the sum of the squared loadings (SSL) of the ASCA model based on the first two components.

Though this method may have sufficed to identify key bins, as is evident from the Venn results reported in the main paper, it was used in conjunction with the cross-sectional (specifically PLS–DA) and RM ANOVA results to ensure we did not exclude potentially important bins. This analysis was performed in Matlab using the statistics toolbox [3] and code provided by Gooitzen Zwanenburg (available under APACHE Licence 2.0 <http://www.apache.org/licenses/LICENSE-2.0.html>).

Section 3.7 Graphs on excretion kinetics

The graphs shown in Figure H provide a visualization of summary statistics for the quantified data of the metabolically important metabolites summarized in Table 1 of the main manuscript. These plots were generated using `notBoxPlot.m` a Matlab function developed by Rob Campbell (<http://www.mathworks.com/matlabcentral/fileexchange/26508-raacampbell13-notboxplot>). The raw data are scattered over a 95% confidence interval (red) and 1 standard deviation of the mean (blue). In two cases, outliers of more than 1.5 times the interquartile range were excluded from the figures and the calculations (Hippuric acid 4250 and creatine of 2500 μmoles metabolite/m mole creatinine, both observed at 1 hour following the intervention).

Reaction kinetics are represented by concentration–time profiles of the six metabolites for the 22 participants that participated in the intervention, shown at times 1, 2, 3 and 4 hours following the intervention (time 0 hour). All quantified values of metabolites in all samples were normalized relative to the creatinine concentration recorded for each sample.

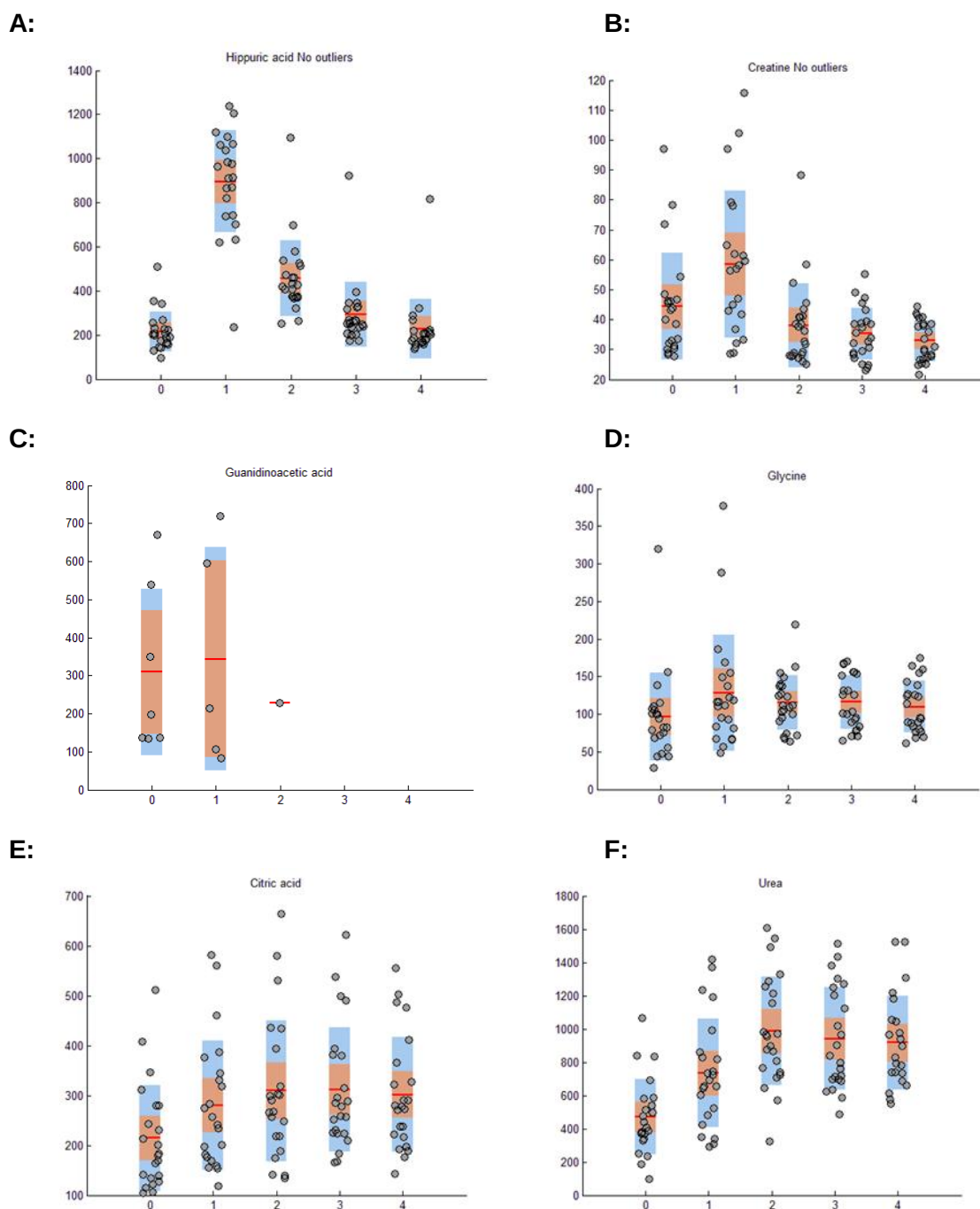


Figure H: Urinary Excretion kinetics of important metabolites

Concentration–time profiles of hippuric acid (A), creatine (B), guanidinoacetic acid (C), glycine (D), citric acid (E) and urea (F) for times 1, 2, 3 and 4 hours following consumption of 500 ml benzoic acid containing flavored water at time 0 hour. All samples contained all metabolites, except for guanidinoacetic acid, as indicated in (C).

Section 3.8 NMR spectra on the excretion of six substances

The authentic spectra and power values from time-dependent PCA's were used to re-analyze the outcomes of the intervention. It appeared that three unknown substances (Figure I: (A) 8.22 ppm, (B) 6.485 ppm and (D) 3.075 ppm), as well as methylguanidine (2.83 ppm - shown in Figure I (C)) — a known endogenous metabolite, changed due to the

intervention. Methylguanidine is synthesized from creatinine concomitant with the synthesis of hydrogen peroxide from endogenous substrates in peroxisomes. It will be speculative to indicate its link to the biotransformation study, but it may be linked to oxidative stress. Although none of the four substances were identified as VIP's in the PLS-DA, their concentrations changed in time following the intervention. As such, they could contribute to the centroid values returning towards the value observed before the intervention, without reaching the exact value of time 0. These observations indicate that the consequence of benzoic acid consumption is more complex than just its biotransformation to hippuric acid.

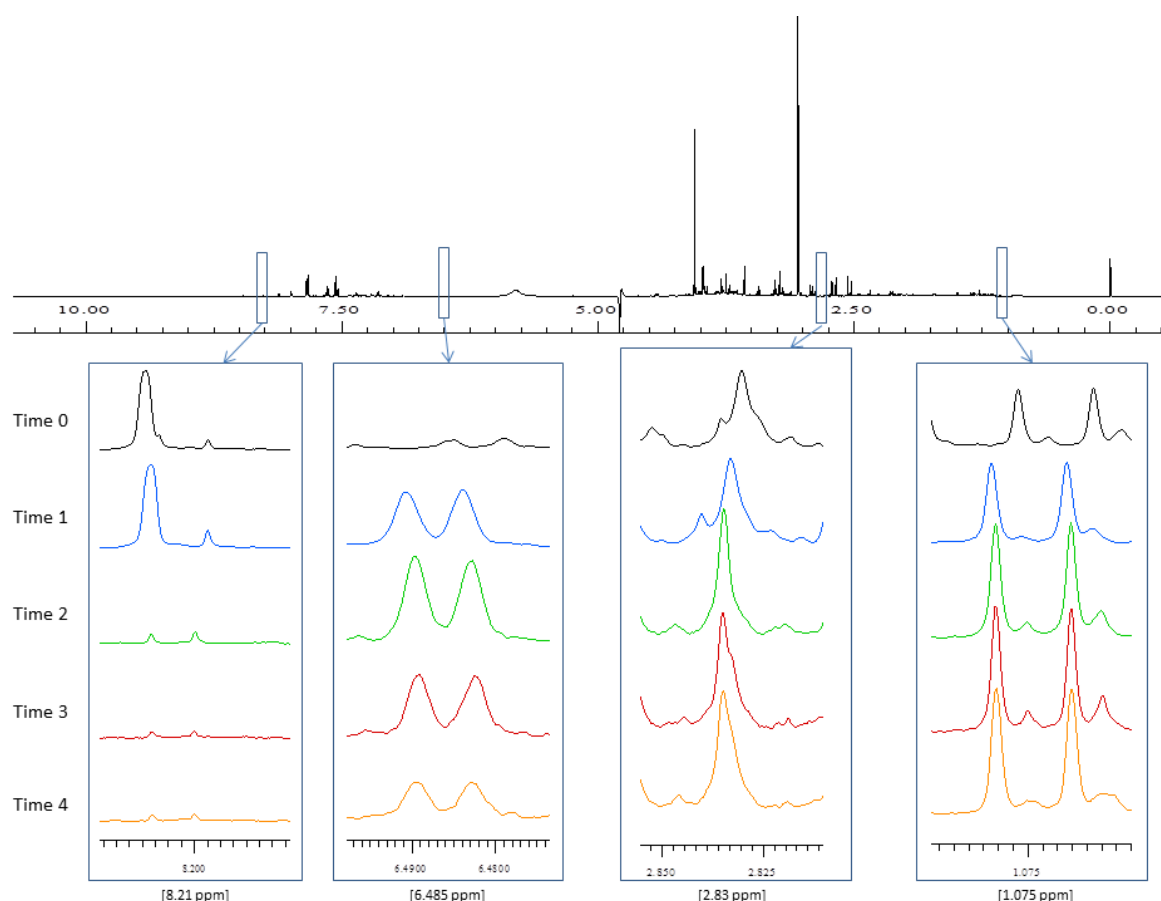


Figure I: 500 MHz ^1H -NMR spectra of minor components from urine. Top spectrum taken from a urine sample of one selected case; spectra from left to right are: (A) unknown (8.11 ppm), (B) unknown (6.485 ppm), (C) methylguanidine (2.83 ppm) and (D) unknown (1.075 ppm). The substances for times 0 (black), 1 (blue), 2 (green), 3 (red) and 4 (orange).

Section 4 NMR analysis on guanidinoacetic acid

Theoretically, the chemical structure of guanidinoacetic acid contains three carbon atoms of which only one has free hydrogen atoms ($-\text{CH}_2-$) that would produce a singlet in the NMR spectrum. This is observed in an example for one of our cases, as shown in the one dimensional (1D) spectrum below. Since guanidinoacetic acid does not have any correlating protons, CORrelation SpectroscopY (COSY) two dimensional analyses is not applicable. In the spectra shown in Figure J we compared the experimental spectrum from one of our samples (in green) to that of the pure guanidinoacetic acid standard (in red) – both in 1D and 2D ^1H J-Resolved (JRES) NMR Spectroscopic analysis. The results indicate that the experimental spectrum lines up exactly to the guanidinoacetic acid singlet (3.77 ppm), confirming the identity of guanidinoacetic acid in the urinary sample.

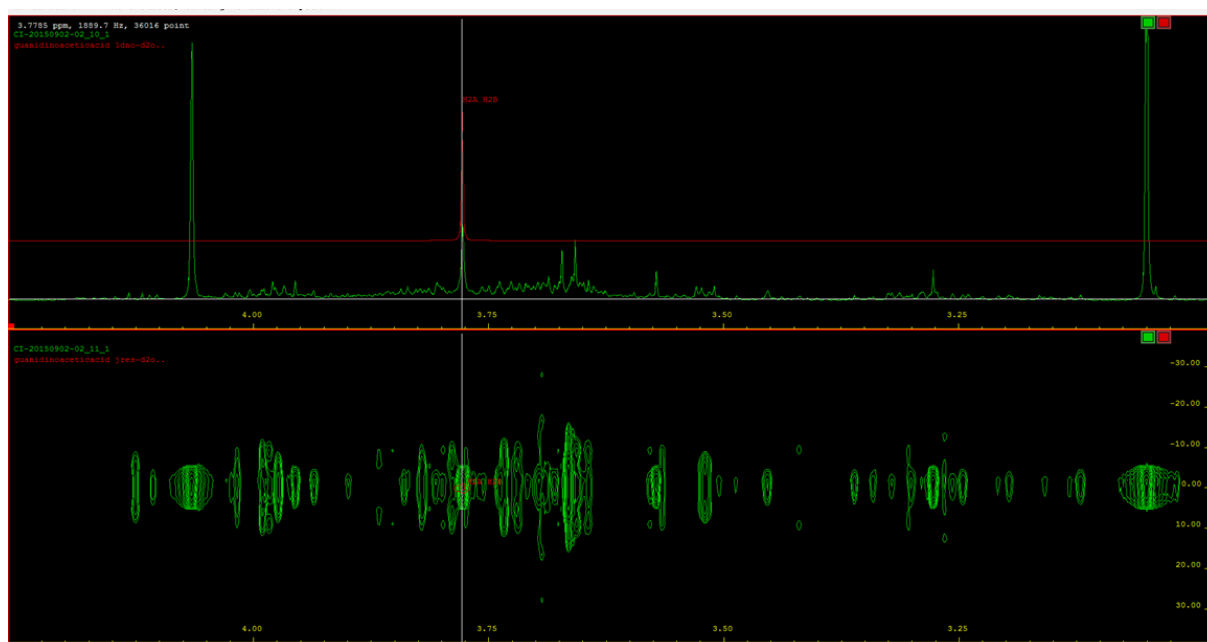


Figure J: 1D and 2D JRES NMR spectra

Analysis of the spectra from a urine sample shown in green and from a guanidinoacetic acid standard shown in red.

References

1. Krojanski HG, Lambert J, Gerikalan Y, Suter D, Hergenröder R. Microslot NMR probe for metabolomics studies. *Anal Chem*. 2008; 80:8668–8672.
2. Smilde AKM, Van der Werf M, Bijlsma S, Van der Werff-van der Vat BJC, Jellema RH. Fusion of mass spectrometry-based metabolomics data. *Anal Chem*. 2005; 77:6729–6736.
3. MATLAB with Statistics Toolbox Release 2012b, The MathWorks Inc., Natick, MA, USA and PLS Toolbox 7.0 (2012), Eigenvector Research.
4. Zwanenburg G, Hoefsloot HCJ, Westerhuis JA, Jansen JJ, Smilde AK. ANOVA–principal component analysis and ANOVA–simultaneous component analysis: a comparison. *J Chemometrics*. 2011; 25:561–567.
5. Viant M, Bearden DW, Bundy JG, Burton IW, Colette TW, Ekman DR, et al. International NMR-based environmental metabolomics intercomparison exercise. *Environ Sci Technol*. 2009; 43:219–225.
6. Van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde KA, Van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*. 2006; 7:1–15.
7. Field AP. *Discovering statistics using IBM SPSS Statistics: and sex and drugs and rock 'n' roll*. 3rd ed. London: Sage Publications; 2009.
8. Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis*. 5th ed. Englewood Cliffs, New Jersey: Prentice Hall; 2002.
9. Westerhuis JA, Hoefsloot HCJ., Smit S, Vis DJ, Smilde AK, Van Velzen EJJ, et al. Assessment of PLSDA cross validation. *Metabolomics*. 2008; 4(1):81–89.
10. Pérez-Enciso M, Tenenhaus M. Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS–DA) approach. *Hum Genet*. 2003; 112:581–592.
11. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 5th ed. Boston: Allyn & Bacon; 2006.
12. Tabachnick BG, Fidell LS. *Experimental Designs using ANOVA*. Belmont: Thomson Brooks/Cole; 2007.
13. Trujillo-Ortiz A, Hernandez-Walls R, Trujillo-Perez RA. RMAOV1: One-way repeated measures ANOVA. A MATLAB file. 2004. <http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=5576>.
14. Villez K, Steppe K, De Pauw DJW. Use of unfold PCA for on-line plant stress monitoring and sensor failure detection. *Biosystems Engineering*. 2009; 103:23–24.
15. Jansen JJ, Hoefsloot HCJ, Van der Greef J, Timmerman ME, Westerhuis JA, Smilde AK. ASCA: analysis of multivariate data obtained from an experimental design. *J Chemometrics*. 2005; 19:469–481.
16. Smilde AK, Jansen JJ, Hoefsloot HCJ, Lamers R-JAN, Van der Greef J, Timmerman ME. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics*. 2005; 21(13):3043–3048.

4.5 An intervention study on acute alcohol consumption

4.5.1 Background

Following the experiment to determine the effect of consuming the vehicle, the subsequent analysis addressed the main purpose of my study — to determine the holistic metabolic consequences of acute alcohol consumption. This part of the study also included examining the potential effect of an NAD supplement on the resulting metabolite profile. To achieve these objectives, this investigation incorporated all four of the interventions referred to in Figure 4–1. The effects of all four interventions were measured over a period of 5 hours, indicated in Figure 1 of Irwin *et al.*, 2016.

The outcome of this investigation, presented in section 4.5.2 as a peer-reviewed paper (accepted for publication in *PLOS ONE*), provided a number of important insights that have the potential to significantly advance our understanding of the impact of the consumption of a single, large dose of alcohol, and include:

1. The metabolite profiles resulting from acute alcohol consumption indicated that alcohol-induced NAD^+ depletion is the primary metabolic consequence of acute alcohol consumption and greatly perturbs the hepatocyte redox homeostasis, resulting in several major metabolic disturbances.
2. The use of NAD as a supplement did not seem to notably counteract the alcohol-induced NAD^+ depletion or the effects thereof.
3. The novel observation of increased urinary excretion of the osmolyte sorbitol is suggested to be linked to hepatocyte swelling due to ethanol influx following acute alcohol consumption.

The combination of these perturbations may underlie the pathophysiology, as well as the range of risks, resulting from and associated with persistent alcohol consumption.

4.5.2 The ¹H-NMR-based metabolite profile of acute alcohol consumption: a metabolomics intervention study (Irwin *et al.*, 2018 – Accepted)

(Accepted for publication in *PLOS ONE*)

Tracking number: PONE-D-17-40411)

The ¹H-NMR-based metabolite profile of acute alcohol consumption: a metabolomics intervention study

Cindy Irwin^{1¶}, Mari van Reenen^{1,2¶}, Shayne Mason¹, Lodewyk J. Mienie¹, Ron A. Wevers³, Johan A. Westerhuis^{2,4} and Carolus J. Reinecke^{1*}

¹Centre for Human Metabolomics, Faculty of Natural Sciences and Agriculture, North-West University (Potchefstroom Campus), Potchefstroom, South Africa

²Department of Statistics, Faculty of Natural Sciences and Agriculture, North-West University (Potchefstroom Campus), Potchefstroom, South Africa

³Department of Laboratory Medicine, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

⁴Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands

*Corresponding author

E-mail: carools.reinecke@nwu.ac.za (CJ)

¶These authors contributed equally to this work

Abstract

Metabolomics studies of disease conditions related to chronic alcohol consumption provide compelling evidence of several perturbed metabolic pathways underlying the pathophysiology of alcoholism. The objective of the present study was to utilize proton nuclear magnetic resonance (^1H -NMR) spectroscopy metabolomics to study the holistic metabolic consequences of acute alcohol consumption in humans. The experimental design was a cross-over intervention study which included a number of substances to be consumed — alcohol, a nicotinamide adenine dinucleotide (NAD) supplement, and a benzoic acid-containing flavoured water vehicle. The experimental subjects — 24 healthy, moderate-drinking young men — each provided six hourly-collected urine samples for analysis. Complete data sets were obtained from 20 of the subjects and used for data generation, analysis and interpretation. The results from the NMR approach produced complex spectral data, which could be resolved sufficiently through the application of a combination of univariate and multivariate methods of statistical analysis. The metabolite profiles resulting from acute alcohol consumption indicated that alcohol-induced NAD⁺ depletion, and the production of an excessive amount of reducing equivalents, greatly perturbed the hepatocyte redox homeostasis, resulting in essentially three major metabolic disturbances — up-regulated lactic acid metabolism, down-regulated purine catabolism and osmoregulation. Of these, the urinary excretion of the osmolyte sorbitol proved to be novel, and suggests hepatocyte swelling due to ethanol influx following acute alcohol consumption. Time-dependent metabolomics investigations, using designed interventions, provide a way of interpreting the variation induced by the different factors of a designed experiment, thereby also giving methodological significance to this study. The outcomes of this approach have the potential to significantly advance our understanding of the serious impact of the pathophysiological perturbations which arise from the consumption of a single, large dose of alcohol — a simulation of a widespread, and mostly naive, social practice.

Keywords: acute alcohol, metabolomics, proton magnetic resonance spectroscopy, metabolite profile, nicotinamide adenine dinucleotide

Introduction

The use of metabolomics in intervention, or challenge, studies greatly enhances the holistic understanding of the effects of single, or combined, consumed substances on metabolic pathways [1,2,3]. In this paper, we present the experimental design and outcome of an intervention study, which included a number of interventions: (1) vehicle only — commercial flavoured water containing sodium benzoate as preservative [3]; (2) a defined acute dose of alcohol, consumed alongside the vehicle; (3) an NAD supplement, taken one hour prior to the study, followed by consumption of the vehicle; and (4) the NAD supplement, followed by consumption of the vehicle and the alcohol dose. Alcohol oxidation by alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH) results in the reduction of oxidized NAD (NAD^+) to its reduced form (NADH), thereby generating a highly reduced cytosolic environment in hepatocytes. An increased NADH: NAD^+ ratio has been described to influence several metabolic processes, resulting in, amongst others, decreased glycolysis [4], reduced Krebs cycle [5,6], and decreased gluconeogenesis [6,7]. The use here of NAD as a supplement is based on a perceived view that it has a therapeutic capacity if taken prior to alcohol consumption; the theory being that it supplements alcohol-induced NAD^+ depletion.

The data set of the present study was generated from longitudinal (urine samples, collected hourly over a five-hour period), multi-subject (several experimental participants) and multi-group (in this case four, separate interventions) measurements, and expressed as multivariate data — numerous variables generated and identified through a metabolomics approach utilizing ^1H -NMR spectroscopy. The main aim of the investigation was to determine the effect of acute (a single dose indigested in a short time) alcohol consumption, as well as the potential effect of an NAD supplement on the resulting metabolite profile. Several metabolomics studies have been reported on the pathophysiological consequences of chronic consumption and alcoholism in humans [8,9,10,11], as well as on chronic intragastric alcohol feeding of rodents, followed by mass spectrometric based metabolite profiling [12,13]. The metabolomics technique has been applied for global metabolite profiling using GC–MS [14,15], in several LC–MS studies [12,13,16,17], as well as in untargeted studies using ^1H -NMR spectroscopy (examples hereof are summarized in Table 1). ^1H -NMR spectroscopy is a robust method with broad metabolite-class coverage (albeit with limited sensitivity), and is well suited for the diverse metabolomics studies listed in Table 1 using rodents [18,19,20,21,22], selected human volunteers at risk for cardiovascular diseases [23], and population-based cohorts [24]. With the exception of the study by Bradford *et al.* [19], all the studies listed in Table 1 used serum or tissue material as the samples for the generation of the metabolite data. Here we report on an untargeted NMR metabolomics intervention study on acute alcohol consumption, which expands the limited data on alcohol-induced metabolic profiling in humans. This study followed a non-invasive sampling procedure (hourly-collected urine samples) which is ideal for

generating the untargeted ^1H -NMR-based metabolic profiles required for the longitudinal experimental design. As a biological waste material, urine is the biofluid of choice [25] to reflect the metabolic breakdown products following acute alcohol consumption, as well as any potential by-products derived from the gut microbiome, which has been shown to be important in several alcohol-related studies [14,16,26]. The data pre-processing applied to the spectral data in the present study are comparable to the approaches applied in the studies listed in Table 1, albeit with minor differences.

Table 1. A summary of the experimental designs, data analysis approaches and main metabolic conclusions of some NMR-based ethanol administration studies.

Authors (year) & Summary of design	Data pre-processing	Statistical analysis	Conclusion on metabolic observations
Nicolas et al. (2008) C: Sprague–Dawley rats (N=5). S: Serum and tissue. E: Single dose/binge protocol. A: Untargeted	(1) Zero-filled. (2) Fourier transformed. (3) Baseline corrected	D: Spectral bins. T: Mean centred. G: Pre vs Post. U: t-test. M: PCA	A cycle of aborted gluconeogenesis stimulated by the increased NADH/NAD ratio but short-circuited by decreased alanine levels during hepatic ethanol metabolism.
Bradford et al. (2008) C: Male C57Bl/6J mice (N=4). S: Urine, serum and liver. E: Chronic: 7–36 g/kg/day; Fed <i>ad libitum</i> . A: Untargeted	(1) Spectra phased. (2) Baseline corrected. (3) Normalized to the sum of all integrals set to 1000	D: Spectral bins. T: Pareto scaled. G: Pre vs Post. U: t-test. M: PCA & OPLS	Energy utilization is important in understanding the pathogenesis of alcohol induced liver injury, with acetylglutamine, n-acetylglycine and taurine potential novel non-invasive markers of alcohol consumption and oxidative stress.
Masuo et al. (2009) C: Female Fisher rats (N=5). S: Liver and brain. E: Chronic: 10 mo 15% Sake; Fed <i>ad libitum</i> . A: Untargeted	(1) Spectra normalize to total integral area. (2) Referenced to Na-TMS-tetra-1H-propionate	D: Spectral bins. T: Pareto scaled. G: Pre vs Post. U: t-test. M: PCA & PLS–DA	In liver: An attenuation of mitochondrial function; In brain: (1) Perturbed amino acids (Increased and decreased); and (2) Decreased N-acetyl-aspartate, taurine and GABA. Chronic Sake intake may cause alterations in the intoxicated body but also in the next generation.
Fernando et al. (2010) C: Male Fischer rats (N=344). S: Plasma and liver. E: Chronic: 5% alcohol diet. A: Lipidome	(1) Spectra manually phased. Referenced to TMS. (2) Spectra divided into equal bins of 0.01 ppm width. (3) Baseline corrected	D: Spectral bins. T: Auto scaled. G: Pre vs Post. U: t-test. M: Hierarchical clustering & PCA	Alcohol consumption alters metabolism of cholesterol, triglycerides and phospholipids that could contribute to the development of fatty liver and indicates that fatty liver precedes oxidative stress and inflammation.
Yoseph et al. (2013) C: Male FVB/N mice (N: 8–9). S: Pancreatic tissue. E: Chronic; Fed <i>ad libitum</i> with increased interventions. A: Untargeted/Focus: Pancreatic metabolome	(1) Spectra phased. (2) Baseline corrected. (3) Spiking experiments to validate suspected metabolites identified in the spectra.	D: 52 Metabolites (mM). T: Variance scaled. G: Pre vs Post. U: t-test & MW test. M: PLS–DA & One-way ANOVA (Tukey post-test)	Pancreatic metabolome following chronic alcohol intake indicates increased acetate, adenosine, xanthine, acetoacetate, 3-hydroxybutyrate and betaine; and (2) decreased cytidine, uracil, fumarate, creatine phosphate creatine, and choline. Mice with chronic alcohol ingestion have increased mortality when encountered with sepsis.
Vázquez-Fresno et al. (2012) C: Males volunteers at risk (Age≥55; N=61). S: 24-hour urine. E: Nutritional wine. A: Untargeted and selected metabolites	Untargeted: (1) Spectra divided into equal bins. (2) Profiling integration. (3) Spectra phased. (4) Baseline corrected. (5) Calibrated. <i>Selected metabolites</i> : Chenomx NMR Suite 7.0 profiler compared on: HMDB/BMRB/MMCD and literature	D: Spectral bins. T: Pareto scaled. G: Pre vs Post. U: t-test; Pearson's correlation. M: ANOVA & Fisher's LSD	(1) Food metabolome: Mannitol (diet); tartrate (wine intake). (2) Endogenous modifications after wine consumption indicated by branch-chain amino acids. (3) Gut metabolites, 4-hydroxy phenylacetate and hippurate
Wurtz et al. (2016) C: 3 population-based cohorts, (meta-analysis N=9778). S: Serum. E: Habitual consumption. A: Targeted lipids and 86 metabolic measures	Automated NMR metabolite profiling: Robotics-controlled and fully automated with a capacity of about 150–180 samples in 24 h. Integrated computational methods for the data-driven systems biology approach to biomedical research	D: Computational generated. T: Log _e transformed & SD Scaled. G: Cross-sectional (i.e. 2 group comparisons). U: Bonferroni method; t-test & MW test. M: Linear regression (R-squared)	Prominent metabolic associations with alcohol consumption include monounsaturated fatty acids, omega-6 fatty acids, glutamine, citrate and lipoprotein particle size. Many of these cardiometabolic biomarkers strongly associated with alcohol intake as HDL cholesterol.

Authors (year) & Summary of design	Data pre-processing	Statistical analysis	Conclusion on metabolic observations
Irwin <i>et al.</i> (the present study) C: Healthy males (Age~22; N=24). S: Hourly urine (5 times). E: Acute (1.5 g vodka/kg). A: Untargeted and quantified metabolites	Normalized relative to the creatinine at 4.05 ppm. Baseline corrected – 50% zero-filter. Batch comparisons using QC samples	D: Spectral bins (excluding H ₂ O region). T: log _e transformed; Auto-scaled. G: Cross-over, time dependent matched series. U: WR-test; ES; FC. M: Hierarchical cluster; Multi-level PCA. The NAD-effect: 2-way RM ANOVA	Aims: (1) Effect of vehicle consumption: Irwin <i>et al.</i> (2016) PLOS ONE. 11:e0167309. (2) Impact of consumption of a single large dose of alcohol (time dependence). (3) Statistical analysis of quantified metabolites following consumption. (4) Effect of NAD supplementation. (5) Model of the metabolite profile following acute alcohol consumption

Abbreviations used in columns 1 and 3: C: Cases studied (experimental animals or humans); S: Samples used for metabolite identification; E: Ethanol exposure; A: Approach for the generation of the variable or metabolite data; N: number of experimental cases studied; D: Data used for statistical analyses; T: Transformation and scaling approach; G: Groups compared; U: Univariate analyses applied; M: Multivariate analyses applied.

Other abbreviations: ES: Effect Size; FC: Fold Change; ANOVA: Analysis of variance; MW: Mann-Whitney test; PCA: Principal Component Analysis; PLS-DA: Partial Least Squares-Discriminant Analysis; RM: Repeated Measures; WC: Wilcoxon test.

The statistical methods used in the intervention studies listed in Table 1 all use a pre vs post approach — multivariate model-based analysis (PCA, PLS–DA, OPLS and ANOVAs) of the data was performed to determine the metabolites responsible for the separation of the control (“pre”) versus the alcohol-treated (“post”) groups. Analysis of the data set from the present study is theoretically complex and required methods specifically designed for longitudinal, multi-subject, multi-group and multivariate data [27,28]. We accordingly generated a matched-sample series through a cross-over study of participating subjects, collecting samples over a distinct time frame. First, data from the “vehicle only” and “alcohol plus vehicle” interventions were compared at each time point to determine the impact of the consumption of a single, large dose of alcohol. Next, the statistical analysis focused on the metabolic perturbations due to the acute consumption of alcohol, measured over time. A shortlist of significantly perturbed NMR spectral bins was compiled and quantified as relative concentrations for biological interpretation. Finally, the effect of NAD supplementation was evaluated for a select list of biologically relevant metabolites.

Most remarkable in the present study was the highly increased presence of urinary sorbitol, a response to alcohol consumption not observed in the studies listed in Table 1, nor previously reported elsewhere. The other main metabolic observations that emanated from this NMR-based intervention study was the increased levels of urinary lactic acid (indicated as a biomarker in two of the listed rodent studies) and hypoxanthine (not observed in the studies listed in Table 1, but observed in another metabolomics study on patients with liver cirrhosis [11]). Both these observations strongly support the view that the reduced cytosolic environment is the primary metabolic consequence of acute alcohol consumption. Taken together, the metabolic information acquired from this metabolomics study further underscores the view that alcohol consumption is associated with severe risks and remains one of the world’s leading health risk factors for disability, morbidity and mortality. This is succinctly expressed in the 2014 World Health Organization global status report on alcohol [29]: *“Of all deaths worldwide, 5.9% are attributable to alcohol consumption; this is greater than, for example, the proportion of deaths from HIV/AIDS (2.8%), violence (0.9%) or tuberculosis (1.7%)”*.

Materials and methods

Chemicals and reagents

The substances used for the interventions reported here were: commercial flavoured water as vehicle (aQuellé lemon-flavoured sparkling water, containing carbonated natural spring water; fructose; citric acid; flavouring; sodium benzoate preservative; sodium cyclamate, aspartame, acesulfame K non-nutritive sweeteners; and vitamin C – www.aquelle.co.za, product of South Africa); commercial alcohol for consumption (Smirnoff No. 21 triple-distilled vodka: 43% alcohol – product of South Africa); NAD (NAD ASSIST – product of Future Health, South Africa); and commercial bottled water (Valpré still spring water, inorganic contents specified – product of South

Africa). The internal standard for the ^1H -NMR analysis was trimethyl-2,2,3,3-tetradeuteriopropionic acid (TSP, sodium salt; Sigma Aldrich).

Experimental subjects and protocol

The participants in the intervention study consisted of a group of 24 medically confirmed healthy males, of various ethnicities, between 20 and 24 years of age. All of the participants were in a healthy, athletic condition, with weights in the 61–92 kg range. They were neither alcohol addicts nor total abstainers, but confirmed their use of alcohol at a moderate social level. No participants took any medication, all were asked to refrain from vitamins, minerals, and other supplementation, and were requested to follow a similar dietary and lifestyle pattern for the duration of the study. The complete protocol, described previously [3], was approved by the Health Sciences Ethical Committee of North-West University (Ethical approval number: NWU-00045-12-S1) and conducted in accordance with guidelines for good clinical practice. The study was performed at the Health Clinic of the university, under the supervision of a medical doctor as well as a nurse, and all participants could leave the premises only after approval by the doctor.

The experiments were conducted on Saturday mornings between 08:00 and 12:00. All participants had to abstain from breakfast and had to provide an early morning urine sample, collected one hour before the start of the experiment (time –1). The participants were randomly assigned to an intervention group until all 24 had participated in all four interventions (“vehicle only”, “alcohol plus vehicle”, “vehicle plus NAD” and “vehicle plus alcohol and NAD”). Owing to commitments of some participants, the experiments were performed over a period of 7 consecutive Saturdays. However, three of the participants failed to participate in all four interventions, and their samples were therefore excluded from the data generation process. Through data analysis, one further participant was marked as an outlier in one of the interventions, and was therefore also excluded from further analyses. Thus, the data generated were based on the complete sets of information obtained from 20 participants.

The four interventions consisted of the consumption of: (1) 500 mL lemon-flavoured water only, to measure the baseline effect of the vehicle; (2) 1.5 g vodka per kg body mass; (3) one tablet containing 50 mg NAD; and (4) 1.5 g vodka per kg body mass plus 50 mg NAD. In all three of the latter interventions, the substances were consumed with 500 mL flavoured water as vehicle. NAD (where applicable) was taken one hour before the start of the experiment, directly after collection of the first (time –1) urine sample. The dose and time schedule for consuming the alcohol (maximally five minutes) comply with established criteria to result in a tolerable but moderately severe level of acute alcohol intoxication [30]. All participants were provided with 1.5 L pure spring water, which was the only substance that could be consumed over the four-hour period of sample collection. Urine samples were collected at time 0, just prior to consumption of the substances, followed by four further samples at 1, 2, 3 and 4 hours thereafter, providing six samples in total from each participant for each of the four interventions. All samples were treated, stored, prepared and analysed according to the protocol described previously [3] and included in Sections A.1 and A.2 in S1 File.

Although repeatability and reproducibility are not major concerns in NMR analyses [31], the measurement design included the use of pooled quality control (QC) samples to estimate any batch effect or other interfering analytical aspect. The collected samples were analyzed in 24 separate batches, each batch containing the 24 samples of a single subject and three QC samples. The batch analysis order is given in Section A.1 in S1 File. The quality assurance results (given in Section A.3.2 and Figure A in S1 File) indicated that the QC samples clustered close together and therefore no batch correction was required.

Uric acid does not have NMR-detectable protons at physiological pH, making it essentially “NMR invisible” [25]. Uric acid concentrations were therefore determined in each sample individually on a ThermoFisher Scientific Clinica Chemistry Analyzer (type 863), using ThermoFisher Scientific Uric Acid Reagents (details on the analysis are included in Section B in S1 File).

Data analysis

The original NMR spectral data (referred to in Section C in S1 File and available in S2 File) were subjected to pre-processing — data were normalized relative to the CH₂ creatinine peak (at 4.05 ppm), and very low values were replaced with zero, before performing a 50% zero-filter (details in Section A.3.1 in S1 File). Further bin reduction, based on batch comparisons of the coefficient of variation of the bins from the QC samples, yielded 289 bins for further analyses. The resulting data were analysed along all dimensions in order to understand the metabolic effect of the different interventions in time, as well as the differences between the participants, even when considering a relatively homogeneous group such as the subjects involved in this study. The analysis was performed on different cross-sections and blocks of the data tensor. Univariate as well as multivariate methods were applied. Data were log transformed (natural log with shift parameter equal to 1) and auto-scaled (per bin) prior to univariate (parametric tests only) and multivariate analysis.

The first analysis aimed to shed light on the impact of the consumption of a single, large dose of alcohol. To illustrate this effect, data from the “vehicle only” and “alcohol plus vehicle” interventions were compared at each point in time. Cross-sectional analysis made use of multivariate techniques, including hierarchal cluster analysis (Euclidian distance and Ward linkage) and multi-level principal component analysis (ML-PCA) [32]. Univariate approaches included determining Wilcoxon signed-rank test (WRT) p-values and associated effect size, as well as fold change (FC) values.

Next, statistical analysis focused on the metabolic perturbations that result from the consumption of alcohol as measured in time. Through the application of a combination of statistical methods, and using time 0 as a reference point, or control group, a shortlist of significantly perturbed NMR spectral bins was compiled and quantified from the “alcohol plus vehicle” intervention data for biological interpretation. This list of important bins was extended through the inclusion of metabolites that were

well presented in the spectra, although they did not contribute significantly to the statistical differentiation within the “alcohol plus vehicle” consumption profiles.

Lastly, the effect of NAD supplementation was of interest. In order to evaluate this intervention, a select list of biologically relevant metabolites was modelled across the time period and across all four interventions using a two-way repeated measures analysis of variance (2-way RM ANOVA) model. The significance of the differences was calculated using Greenhouse–Geisser-corrected p-values. In addition, Wilcoxon signed-rank tests were performed to calculate the significance of the differences between selected pairs of measurements.

Results

An example of a ^1H -NMR spectrum (10.00 to 0.00 ppm), generated from the analysis of a representative experimental sample collected one hour following the “alcohol plus vehicle” intervention, is shown in Fig 1a, together with enlargements of four sections of the spectrum to further illustrate some detail (Fig 1b–e).

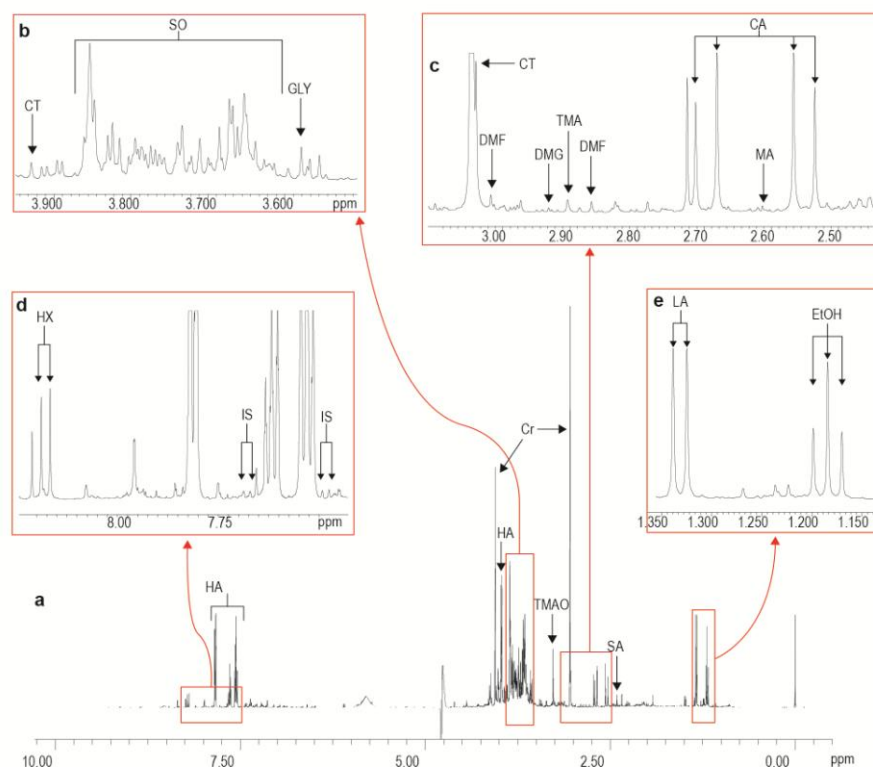


Fig 1. Representative ^1H -NMR spectrum of urine collected one hour following the “alcohol plus vehicle” intervention. EtOH = ethanol (1.18 t, 3.64 q); LA = lactic acid (1.33 d, 4.12 q); SA = succinic acid (2.41 s); TMAO = trimethylamine-N-oxide (3.27 s); CA = citric acid (2.61 AB); MA = methylamine (2.61 s); TMA = trimethylamine (2.90 s); DMG = N,N-dimethylglycine (2.93 s); DMF = N,N-dimethylformamide (2.87 s, 3.02 s); CT = creatine (3.04 s, 3.93 s); GLY = glycine (3.57 s); SO = sorbitol (3.60–3.69 m, 3.73 d, 3.74–3.80 m, 3.82 d, 3.85 m); IS = indoxyl sulphate (7.51 d, 7.70 d); HX = hypoxanthine (8.20 d); HA = hippuric acid (3.97 d, 7.56 tt, 7.64 tt, 7.84 dd); Cr = creatinine (3.05 s, 4.06 s). [Not observed in the present spectrum: fumaric acid (6.52 s) and 3-hydroxybutyric acid (1.20 d, 2.36 m, 4.15 m)].

The dominating peaks in Fig 1a were from ethanol (due to the alcohol consumption), hippuric acid (the biotransformation product of benzoic acid, derived from the gut microbiome and present in the vehicle used in all the interventions, as well as observed in a moderate red wine nutritional study) [23], creatinine (a normal constituent of urine), and trimethylamine-N-oxide (TMAO), a known osmolyte and protein stabilizer. Most notable was the presence of an exceedingly complex area, approximately between 3.60 and 3.90 ppm (Fig 1b). Interpretation of this region is particularly difficult, since it may contain overlapping resonances from several metabolites. The strong signals observed at 3.60–3.69, 3.73, 3.74–3.80 and 3.85 ppm were suggestive of sorbitol in the urine, following alcohol consumption. This suggestion was confirmed by comparing the spectra obtained through two-dimensional (2D) correlation spectroscopy (COSY) analysis of an experimental sample, obtained one hour after alcohol consumption, and a sample containing sorbitol as a standard (Fig 2).

Next, an area with several related amines or amine derivatives is shown in Fig 1c, while Fig 1d indicates the presence of hypoxanthine (an intermediate in the catabolism of purines towards uric acid) and indoxyl sulphate (a biotransformation product of tryptophan). Finally, Fig 1e indicates the presence of ethanol itself, as well as the lactic acid doublet (1.33 ppm), the main known marker of lactic acidosis, and one of the primary effects of acute alcohol consumption.

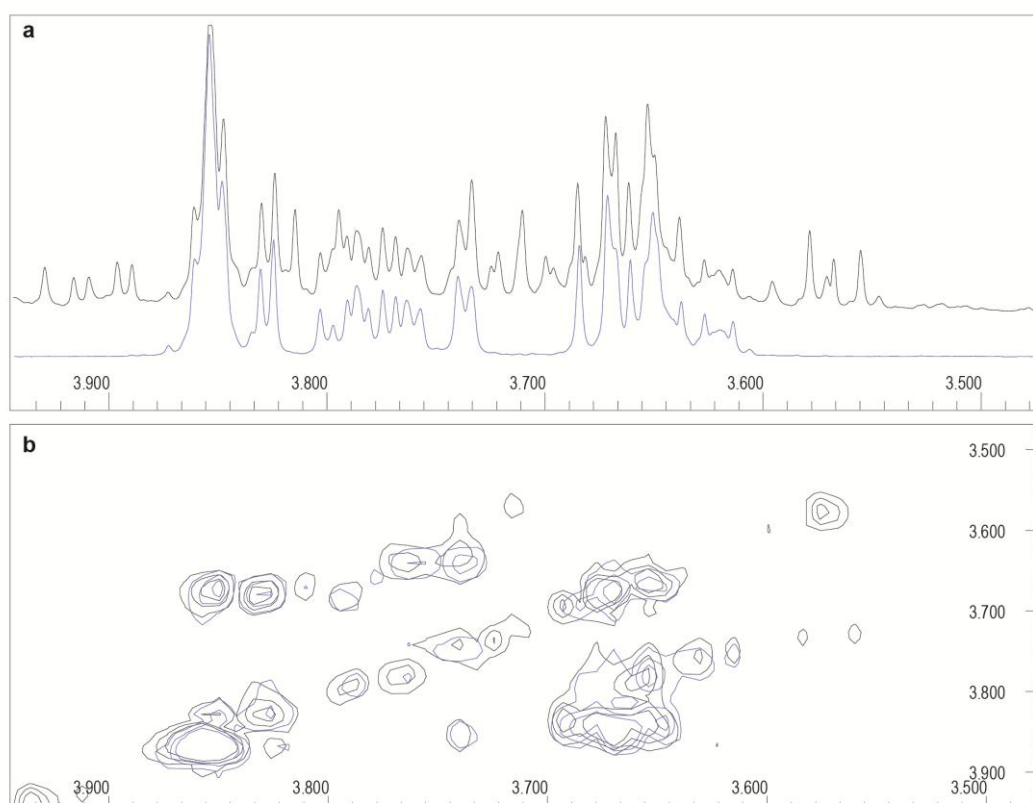


Fig 2. Confirmation of sorbitol annotation. (a) 1D ^1H -NMR of a representative urine sample collected one hour after alcohol consumption, zoomed into the 3.50–3.90 ppm region (black), compared to the pure compound spectrum of sorbitol (blue). (b) Correlating 2D ^1H -NMR COSY, confirming the sorbitol annotation based upon proton correlation.

Spectral analysis of the representative sample gives a clear indication of the important metabolites present in urine following alcohol consumption (Fig 1), but no information on the underlying dynamic effects following the consumption. The next step was therefore to determine the time-dependent impact of the consumption of a single, large dose of alcohol. Statistical analyses of the data obtained from the experimental samples from the “vehicle only” and “alcohol plus vehicle” interventions provide such information, and are illustrated in Fig 3. The NMR spectra generated were dominated by the presence of hippuric acid (see [3]). Exclusion of hippuric acid from the data set, however, did not significantly affect the relevant information on alcohol consumption.

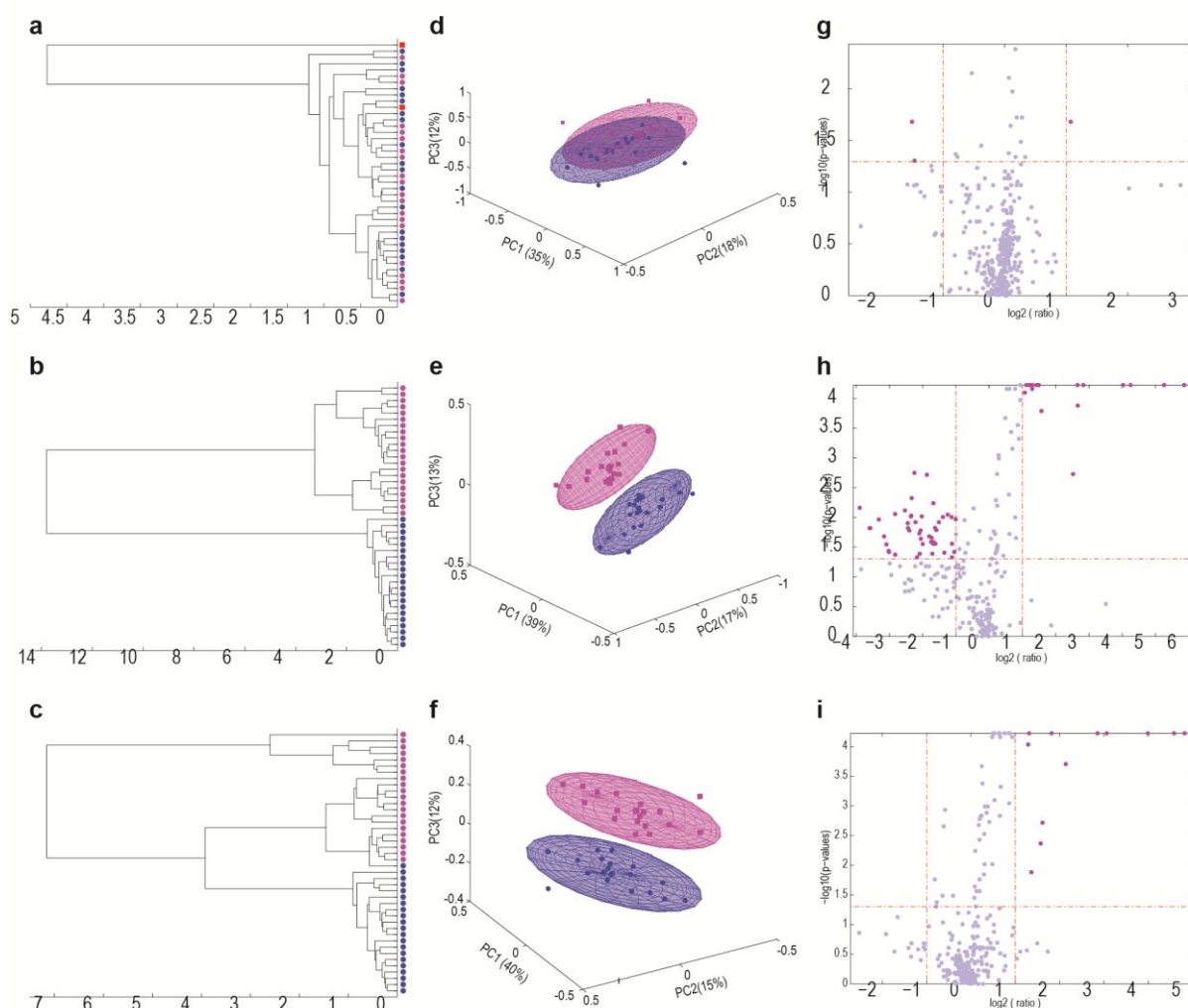


Fig 3. Group separation between participants, based on equidistant binned spectral data from the “vehicle only” and the “alcohol plus vehicle” interventions, illustrated as dendrograms, ML-PCA plots and Volcano plots. The respective analyses were constructed on subsets of the data representing the same three time points — time 0 (a, d and g), 2 hours (b, e and h) and 4 hours (c, f and i) following the two interventions. Data from the 21 participants in the dendrograms and ML-PCA plots are shown as blue dots/areas for the “vehicle only” intervention and pink dots/areas for the “alcohol plus vehicle” intervention. The single outlier is shown as a red square in the dendrograms. All data from this participant were excluded from further analyses, resulting in the analysis of the data from a total of 20 participants.

Figs 3a and 3d (time 0) showed no group separation prior to vehicle or alcohol consumption. It did, however, become apparent that one individual may have used an exogenous substance on the day of his “vehicle only” intervention experiment, which completely separated this person from all others in the group (indicated as a red square in Fig 3a). This rendered the data from this individual unfit for analysis, producing the final number of 20 cases used for further analysis. The bin profiles obtained two hours after the “vehicle only” and “alcohol plus vehicle” interventions clearly indicated group differences in the unsupervised analyses due to the addition of alcohol (Figs 3b and 3e). The Volcano plot from data collected one hour after alcohol consumption revealed a larger number of bins with significant up- and down-regulated values ($p \leq 0.05$ and $|FC| \geq 2$) relative to the number of bins from data collected one hour after vehicle consumption (data not shown). This number of significant bins increased to at least 58 out of the 289 bins (20%) two hours following consumption (Fig 3h). The number of these bins progressively decreased 3 and 4 hours after consumption, becoming only 13 bins after 4 hours that differed significantly between the two interventions (Fig 3i).

Given this marked effect, the next set of analyses were performed to identify and rank the bins most affected by the consumption of alcohol using time 0 as a point of reference. Bins were shortlisted if they differed significantly within the “alcohol plus vehicle” intervention at any point in time (relative to time 0) based on a significant WRT p -value ≤ 0.05 and $|FC| \geq 2$. This shortlist of bins was linked to a set of metabolites and quantified. This list, including only the most perturbed metabolites, was then extended through the inclusion of metabolites which were well presented in the spectra, but did not contribute significantly to the statistical differentiation within the “alcohol plus vehicle” consumption profiles. Quantified uric acid data (not identified by NMR analysis, but determined individually for each urine sample) was also added to this final shortlist of 13 important metabolites. The concentrations of these 14 metabolites, at all five time points related to the “alcohol plus vehicle” intervention, are given in Table 2, together with the relevant summary statistics, determined one (early effect) and four (late effect) hours after alcohol consumption. Time -1 is not included in Table 2 since it reflects the past several hours prior to the intervention, and is not related to the “alcohol plus vehicle” intake. The purpose of Table 2 was then to rank metabolites for biological interpretation of the consequences of the intervention.

Table 2. Quantified data of important metabolites following alcohol consumption.

Variable	Time 0 vs Time 1 (early effect)		Time 0	Time 1	Time 2	Time 3	Time 4	Time 0 vs Time 4 (late effect)	
	WRT p-value	WRT Effect Size	Mean	Mean	Mean	Mean	Mean	WRT p-value	WRT Effect Size
	[BH Adjusted p-value]	[FC]	[SD]	[SD]	[SD]	[SD]	[SD]	[BH Adjusted p-value]	[FC]
DMF (N,N-dimethylformamide)	0.057 [0.079]	0.301 [+1.038]	15.98 [2.292]	16.58 [2.613]	15.64 [2.795]	16.01 [2.308]	17.12 [2.699]	0.017 [0.047]	0.378 [+1.071]
DMG (N,N-dimethylglycine)	0.052 [0.079]	0.307 [+1.140]	8.472 [7.965]	9.656 [10.02]	9.190 [9.230]	8.168 [7.640]	8.438 [7.274]	1.000 [1.000]	0.000 [−1.004]
Ethanol	<0.0001 [0.0004]	0.620 [> +100]	0.000 [0.000]	1389 [1556]	6448 [2673]	3440 [1984]	1912 [1250]	<0.0001 [0.001]	0.620 [> +100]
Glycine	0.044 [0.077]	0.319 [+1.145]	160.2 [84.90]	183.4 [86.26]	151.4 [44.68]	146.6 [50.45]	158.3 [56.11]	0.601 [0.701]	0.083 [−1.012]
Hippuric acid	0.313 [0.313]	0.159 [+1.725]	220.4 [217.2]	380.2 [393.7]	87.66 [124.4]	138.1 [344.7]	178.9 [333.7]	0.100 [0.156]	0.260 [−1.232]
Hypoxanthine	<0.0001 [0.0004]	0.620 [+8.187]	9.604 [7.029]	78.63 [45.97]	29.82 [14.99]	14.44 [5.716]	12.48 [5.191]	0.044 [0.077]	0.319 [+1.299]
Indoxyl sulphate	0.002 [0.005]	0.490 [+1.147]	24.30 [10.29]	27.86 [13.42]	24.87 [16.04]	21.51 [14.25]	22.87 [11.19]	0.794 [0.855]	0.041 [−1.063]
Lactic acid	0.001 [0.004]	0.508 [+2.537]	70.90 [61.39]	179.9 [274.8]	92.74 [19.59]	75.96 [14.04]	72.79 [16.26]	0.006 [0.022]	0.431 [+1.027]
Methylamine	0.006 [0.013]	0.431 [+1.159]	4.556 [2.146]	5.282 [2.477]	5.937 [2.256]	5.186 [2.380]	5.059 [2.263]	0.033 [0.073]	0.336 [+1.110]
Sorbitol	<0.0001 [0.0004]	0.620 [> +100]	0.000 [0.000]	653.2 [462.0]	852.8 [625.7]	426.2 [331.6]	289.0 [253.3]	0.000 [0.001]	0.614 [> +100]
Taurine	0.062 [0.079]	0.295 [+1.076]	100.6 [29.71]	108.2 [32.54]	109.7 [29.86]	113.5 [37.45]	118.7 [36.49]	0.006 [0.022]	0.437 [+1.180]
TMAO (trimethylamine N-oxide)	0.002 [0.004]	0.502 [+1.139]	51.18 [23.90]	58.32 [25.48]	60.31 [29.26]	62.45 [35.68]	62.00 [32.48]	0.037 [0.073]	0.331 [+1.211]
Trimethylamine	0.100 [0.117]	0.260 [+1.133]	1.753 [1.229]	1.987 [1.657]	1.763 [1.680]	1.654 [1.578]	1.615 [1.436]	0.167 [0.234]	0.218 [−1.085]
Uric acid	0.247 [0.266]	0.183 [−1.351]	0.963 [0.760]	0.713 [0.279]	0.336 [0.168]	0.487 [0.243]	0.743 [0.566]	0.211 [0.269]	0.198 [−1.296]

All quantified values, except those for uric acid (expressed as mmol/L), are from NMR-determined urine analyses, and are expressed as μmol metabolite/mmol creatinine. WRT p-values are based on the comparison of the respective metabolite concentrations relative to time 0 for time 1 and time 4 of the “alcohol plus vehicle” intervention. P-values adjusted for multiple testing (14 tests in total) are also reported based on the Benjamini & Hochberg (BH) approach for controlling the rate of false discoveries. Positive and negative fold change values indicate up- and down-regulation of metabolites, relative to time 0, respectively.

The data summarized in Table 2 indicate that seven metabolites (ethanol, hypoxanthine, indoxyl sulphate, lactic acid, methylamine, sorbitol and TMAO) were significantly up-regulated ($p \leq 0.05$) in the first hour following alcohol consumption. Five of these metabolites (ethanol, hypoxanthine, lactic acid, sorbitol and TMAO) remained up-regulated at every time point up to time 4 after alcohol consumption, although less significantly so ($p \leq 0.05$) than after the first hour. This observation illustrates a general characteristic of metabolic profiling, well known in the area of inherited diseases: serial examinations of urinary metabolites show that the amounts of these acids excreted varies greatly in time following a perturbation, mostly due to different metabolic consequences related to the perturbation. The uric acid excretion profile observed in this study further emphasizes this characteristic. The profile of urinary uric acid differed distinctly from those of the other metabolites mentioned — its concentration decreased from time 0, and became significantly reduced 2 and 3 hours after alcohol consumption ($p \leq 0.05$ relative to time 0). Thereafter, its concentration steadily increased to near its initial level after 4 hours. Three other metabolites, DMF, taurine and glycine, became significantly up-regulated at some stage following alcohol consumption. The up-regulation of two metabolites, DMG and trimethylamine, was not significant at any time point relative to time 0 following the intervention. The excretion profile of hippuric acid followed the same trend as previously reported for the “vehicle only” intervention [3].

These mean values of the metabolites over time indicate the dynamic aspect following alcohol consumption, but do not reflect individual responses towards the intervention. Inter-individual variation has already been well established even for consumption of only the vehicle used in this study [3], and therefore, this aspect is not discussed in detail here. Instead, focus was placed on the individual, as well as the group, concentration changes of the five important metabolites that were significantly up-regulated during the whole time period following alcohol consumption, as well as on the concentration changes of uric acid, owing to its unique excretion profile.

Fig 4 displays subplots as a set of solid lines (and one dotted line) representing the observations from the 20 individuals across 5 time points for each of the six metabolites shown to be significantly up-regulated in Table 2. The black dashed lines represent one potential LOWESS (locally weighted scatterplot smoothing) regression for each metabolite against time using a bi-weight kernel. Most individuals excreted a limited amount of ethanol in their urine one hour after alcohol consumption, and, generally, ethanol excretion peaked two hours following alcohol consumption (Fig 4a). Likewise, most individuals showed a distinct excretion of the osmolyte sorbitol one hour after alcohol consumption (Fig 4b), albeit at different concentrations. The LOWESS regression curve for sorbitol, however, peaked at 2 hours for the group as a whole. The concentration of TMAO, another osmolyte, increased significantly from time 0 (Fig 4c), reaching a plateau at 3 hours. Its final concentration was, however, still higher ($p = 0.037$) than the baseline value at time 0.

The next observations from Fig 4 are closely associated with the known alcohol-induced disturbance of the $\text{NAD}^+:\text{NADH}$ ratio in hepatocytes. The excretion of lactic acid for the group peaked at 2 hours, but 4 individuals showed an early and excessive response one hour after alcohol consumption (Fig 4d). The profile of the LOWESS regression curve for hypoxanthine (Fig 4e) distinctly peaked one hour following alcohol consumption. Two of the early high lactic acid responders showed similar, high hypoxanthine excretion profiles. The general profile of uric acid excretion (Fig 4f) was a mirror image of that of hypoxanthine, indicating the NAD^+ -dependence of its formation through dehydrogenation of hypoxanthine (catalysed by xanthine dehydrogenase; EC 1.17.1.4).

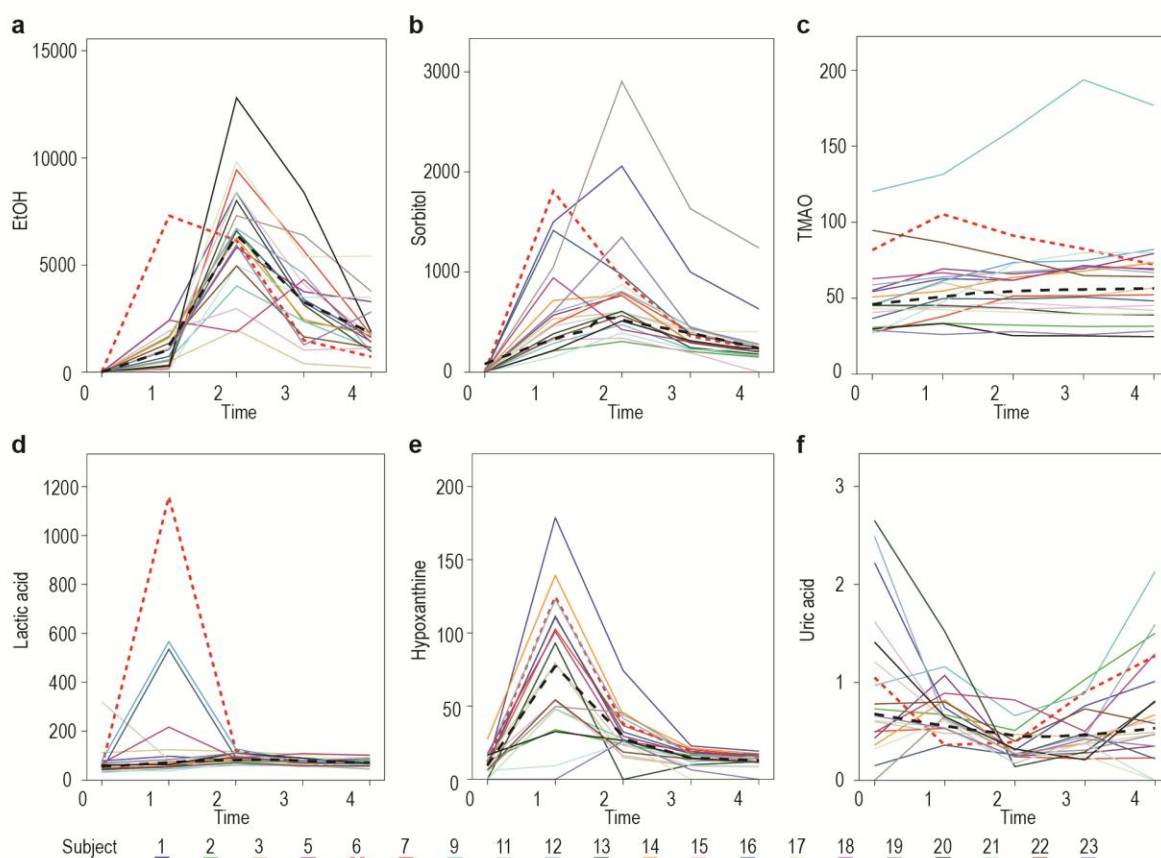


Fig 4. Changes in the concentrations of the six up-regulated metabolites from time 0 to time 4 following alcohol consumption. Each subplot displays a set of lines (coloured solid lines for 19 of the individuals and a red dotted line for one highlighted individual) representing the observations from the 20 individuals across 5 time points for a given metabolite: (a) ethanol; (b) sorbitol; (c) TMAO; (d) lactic acid; (e) hypoxanthine; and (f) uric acid. The black dashed lines represent one potential LOWESS regression for each metabolite against time.

Most individuals followed the general trend indicated by the LOWESS regression line, although one showed an early excretion of ethanol following alcohol consumption (shown as a red dotted line in Fig 4). Compared to the group, this individual also presented with a high excretion of sorbitol and TMAO, peaking at one hour, an early and extremely

high lactic acid excretion, high excretion of hypoxanthine, but comparable excretion of uric acid. Since this person did not present as an outlier, we attribute this variation to the unique response of this individual to alcohol consumption. Diversity in the excretion profiles of the individuals within the group, emphasized by the one highlighted person, clearly illustrates the importance of inter-individuality, previously described for the consumption of only the vehicle [3]. Taken together, the combination of observations summarized in Table 2 and illustrated in Fig 4 indicate that the cases themselves are a noteworthy source of variation, while the group response provided the opportunity for the development of a holistic model of the effect of acute alcohol consumption.

Finally, the ingestion of one tablet of an NAD-containing supplement, proposed to potentially counteract the effects of alcohol consumption, did not show a notable effect if taken one hour before the alcohol dose. To illustrate this, the complete design of this study (that is, four interventions measured over 6 time points) was modelled for quantified hypoxanthine and sorbitol, using a 2-way RM ANOVA model based on log transformed data to improve normality. Reported in the graphs are the Greenhouse–Geisser-corrected p-values for the main effects (where significant), as well as specific comparisons based on the Wilcoxon signed-rank tests (Section A.3.3 in S1 File provides more details on this analysis). These two metabolites were selected based on the NAD^+ -dependence of their catabolism (hypoxanthine $\rightarrow$ xanthine $\rightarrow$ uric acid, both reactions catalysed by xanthine dehydrogenase; and sorbitol $\rightarrow$ fructose, catalysed by sorbitol dehydrogenase (SD; EC 1.1.1.14), respectively). Fig 5 provides a brief overview of the information extracted from this model.

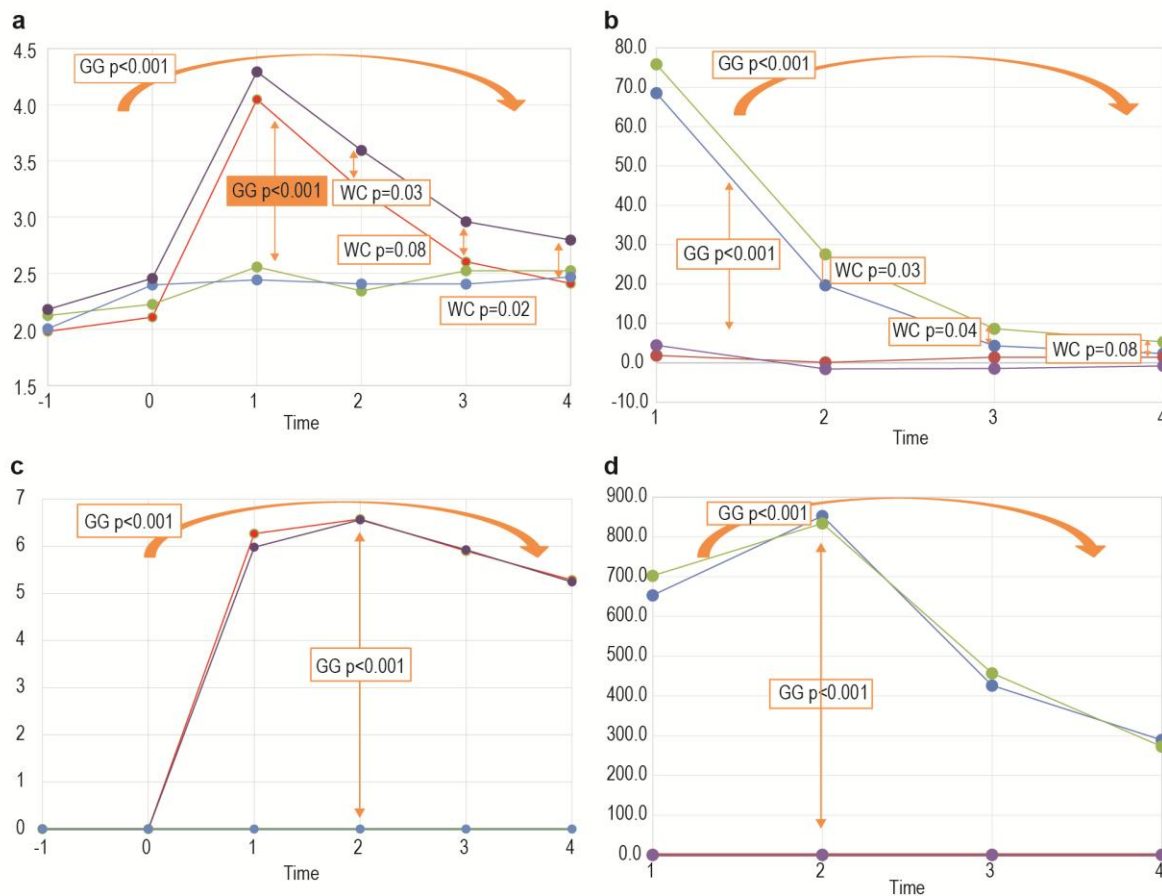


Fig 5. Indications of differences in the average levels of hypoxanthine and sorbitol across the four interventions and six time points. (a) Differences in the average levels of hypoxanthine from time -1 to time 4 between the four interventions. (b) Differences in the average change in hypoxanthine levels, measured from time 0, following the four interventions. (c and d) The comparative results for sorbitol. Significant differences were based on the Greenhouse–Geisser-corrected p-values from the RM ANOVA model or the Wilcoxon signed-rank tests, assessing differences between the sets of means, and are indicated by the arrows.

Figs 5a and 5c indicate a significant difference in the average levels of hypoxanthine and sorbitol, respectively, between the interventions including and excluding alcohol, as well as in time between the interventions including alcohol. Figs 5b and 5d, likewise, indicate a significant difference in the average change in the levels of hypoxanthine and sorbitol across the time period between the four interventions. All the results from this approach therefore arrive at the same conclusion — the observed differences were mainly attributable to the consumption of alcohol. For sorbitol, the differences were only significant when comparing the two interventions involving alcohol with the two not involving alcohol. No significant differences were observed between the two interventions not involving alcohol, nor between the two interventions involving alcohol. For hypoxanthine, significant differences were observed between the “alcohol plus vehicle” and the “vehicle plus alcohol and NAD” interventions at 2 and 3 hours after consumption, with significance on the

average levels at 2 and 4 hours, respectively. Although alcohol consumption caused increased hypoxanthine excretion, it appears that the addition of NAD resulted in even higher amounts of hypoxanthine being excreted. These results indicate that NAD, as used in this study, might potentially have an effect in the later phases following alcohol consumption, but, as a whole, do not justify its application as a supplement to counteract the effects of alcohol consumption.

Discussion

It is known that three mechanisms for ethanol metabolism may be operative in the liver: (1) the ADH-pathway in the cytosol; (2) the microsomal ethanol oxidizing system (MEOS) of the endoplasmic reticulum; and (3) the catalase mechanism, located in peroxisomes [33]. In the ADH-pathway, ethanol is converted to acetaldehyde and acetate in two consecutive dehydrogenase reactions, both dependent on NAD^+ . During both of these reactions NAD^+ is reduced, generating excess amounts of NADH. Following acute alcohol consumption, NAD^+ depletion, and the subsequent excessive production of reducing equivalents, greatly perturb the hepatocyte redox homeostasis. This perturbation is known, but, through the present NMR metabolomics study, new insights were revealed — illustrated in Fig 6. This figure proposes a model that outlines three essential disturbances that occur following acute alcohol consumption — up-regulated lactic acid metabolism, down-regulated purine catabolism and osmoregulation.

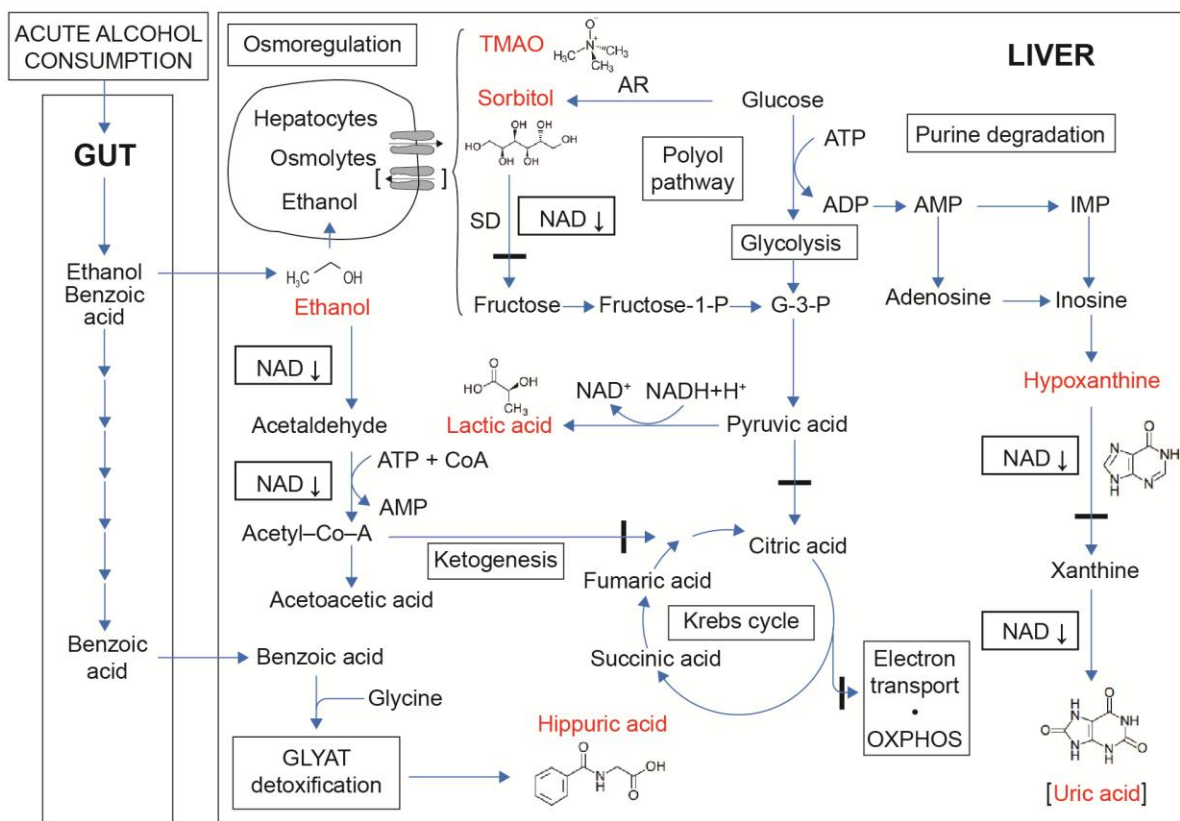


Fig 6. Model of the metabolite profile based on the important metabolites up-regulated following alcohol consumption. The main organs involved are the gut and the liver. Ethanol absorption is indicated in the upper region of the gut, and benzoic acid (from the vehicle and microbiome metabolism) in the lower gut. The main metabolism of ethanol, and the associated consequences of the disturbed $NAD^+ : NADH$ ratio, occur in the liver. The structural formulas and names of the six important metabolites are shown in red (uric acid is indicated in brackets because it was not measured by NMR). Blue arrows are not related to enzyme kinetic reactions, but are used as indicators of the proposed flow directions following alcohol consumption. Osmoregulation is proposed as efflux in hepatocytes in the early phases following alcohol consumption, and potential influx (shown in brackets) in the later phases. Abbreviations: ATP, adenosine triphosphate; ADP, adenosine diphosphate; AMP, adenosine monophosphate; IMP, inosine monophosphate; G-3-P, glyceraldehyde-3-phosphate; GLYAT, glycine-N-acyltransferase; SD, sorbitol dehydrogenase; AR, aldose reductase; OXPHOS, oxidative phosphorylation.

First, up-regulation of lactic acid is coupled to the high hepatic $NADH : NAD^+$ ratio (due to ADH- and ALDH-catalysed ethanol catabolism), which diverts pyruvic acid metabolism towards lactic acid, and subsequently inhibits gluconeogenesis. During low or chronic alcohol consumption, excess lactic acid is exported from the liver to peripheral tissues, where $NADH$ levels are lower, and lactic acid may be reconverted to pyruvic acid for metabolic needs. During acute alcohol consumption, however, aerobic oxidation (that is, the Krebs cycle, the respiratory chain and oxidative phosphorylation) is known to be inhibited [5,6], which was indicated in our study by the increased levels of fumaric acid (not

observed in all experimental subjects) and succinic acid. Similarly, due to the high hepatic NADH:NAD⁺ ratio and inhibition of the Krebs cycle, ethanol-derived acetyl-CoA may be converted to acetoacetic acid and 3-ketobutyric acid, as seen in the urine samples from some of our experimental subjects. The values of these metabolites were, however, generally too low to enable their quantification. A noteworthy observation was that a few of the experimental subjects showed excessive urinary lactic acid excretion shortly after alcohol intake (Fig 4d), whereas the excretion of the majority of the subjects peaked at two hours following alcohol consumption. These observations are yet a further example of the individual differences in coping with the consumed alcohol. The common paradigm is that variation in response to alcohol consumption is genetically controlled, and is suspected to cause a predisposition towards the development of alcohol-induced liver disease and alcoholism [34]. Furthermore, we observed that the mean value of urinary lactic acid peaked one hour after alcohol consumption (increasing roughly 2.5 times from 70.90 mM to 179.9 mM ($p = 0.001$)) and declined to a near normal value three hours after consumption (75.90 mM). This is comparable to the decreased lactic acid levels measured in liver and serum samples from rats, decapitated three hours following treatment with a single intragastric dose of ethanol [18].

Second, hypoxanthine, an intermediate in purine catabolism and a precursor of uric acid, appeared to be an important indicator of acute alcohol consumption (Fig 4e). This is in agreement with the increased hypoxanthine observed in patients suffering from alcohol- and hepatitis B-induced cirrhosis [11], and in those listed with stearamide as biomarker for hepatic cirrhosis [17]. In addition, several intravenous ethanol infusion, and related, studies on purine metabolism in humans (reviewed in [35]) indicated that many factors affect this metabolic pathway — daily drinking habits; the type of alcoholic beverages; exercise; and, ultimately, ALDH polymorphisms, are all important contributing factors. Increased AMP following alcohol consumption seems to be the important metabolic departure point in reflecting on purine catabolism — alcohol-induced diminished ATP production via glycolysis [36] and increased adenine nucleotide turnover [37], contribute towards increased ADP levels, and its increased conversion to AMP. Additionally, during the ADH- and ALDH-catalysed degradation of ethanol, two equivalents of ATP are consumed and two equivalents of AMP are produced for each equivalent of ethanol converted to acetyl-CoA [38]. All these perturbations lead to a considerable alcohol-induced increase in AMP, which is then catabolised and reflected in the hypoxanthine and, ultimately, uric acid profiles. Hyperuricaemia, gout and increased urinary uric acid excretion, accompanied by raised urinary hypoxanthine and xanthine concentrations, are accepted indicators of chronic alcohol consumption, although excretion of hypoxanthine was shown to be lower than that of xanthine in regular drinkers [39]. In our study, we observed a highly significant increase (Table 2: $p \leq 0.0001$; FC = +8.187) in hypoxanthine excretion in all 20 cases studies, which peaked at one hour after alcohol consumption in 18 of the cases, and at 2 hours in the other two cases (Fig 4e). The general pattern of the urinary excretion of uric acid (Fig 4f) presents as a mirror image to that of hypoxanthine, initially decreasing and

then returning to near its starting concentration. We relate these observations to the alcohol-induced perturbed hepatic redox state — depleted NAD^+ , the cofactor for xanthine dehydrogenase, interrupts the purine catabolic pathway, decreasing the conversions of hypoxanthine to xanthine, and xanthine to uric acid. However, as the $\text{NADH}:\text{NAD}^+$ ratio normalizes, these reactions can once again take place, resulting in the catabolism of hypoxanthine to produce uric acid. It should, however, be noted that lactic acid, which is highly increased following alcohol consumption, competitively inhibits the clearance of uric acid through the renal proximal tubule [40], which further supports the observed initial decrease in uric acid excretion following alcohol consumption. Thus, while hyperuricaemia is an accepted marker of regular and chronic alcohol consumption, increased excretion of hypoxanthine may possibly act as an indicator of acute alcohol consumption.

Third, of notable interest, was the observation of the increased urinary excretion of sorbitol following ethanol consumption, which is not listed in the review on global metabolic profiling studies on alcohol-related disorders, covering NMR, GC–MS and LC–MS approaches [17]. Sorbitol was not detected in any of the urine samples collected prior to any of the interventions (normal reference value: $3.506 \pm 2.24 \mu\text{mol}/\text{mmol creatinine}$) [25]. Its excretion, however, increased rapidly in most individuals following alcohol consumption (Fig 4b). Moreover, the sorbitol excretion profile did not seem to be affected by the consumption of the NAD supplement, taken prior to alcohol consumption (Figs 5c and 5d). Sorbitol is an organic osmolyte present in all human cells, and, together with other osmolytes, reaches very high concentrations (in the millimolar range) in the cytosol. In human cells, the osmolytes are classified into three groups [41]: (1) amino acids and their derivatives (including taurine); (2) methylamines (including TMAO); and (3) polyols (including sorbitol). Osmolytes play key roles as cytoprotectants, and in maintaining cell volume homeostasis [42]. They function as nonperturbing solutes, which permits their accumulation to high levels and large shifts in their concentrations without having deleterious effects on cellular structure and function [41]. These unique characteristics of osmolytes opened several lines of thought regarding the perturbation of sorbitol following acute alcohol consumption. Of these postulates, sorbitol synthesis and catabolism (occurring in the polyol pathway), as well as its function in osmoregulation, are pertinent:

(1) The NADPH-dependent enzyme aldose reductase (AR; EC 1.1.1.21) catalyses the synthesis of sorbitol from glucose. This reaction is highly operative under hyperglycaemic conditions, such as in diabetes mellitus [43], when up to 30% of glucose is channelled into the AR-catalysed polyol pathway. However, given the fasting state of the experimental subjects before the “alcohol plus vehicle” intervention, and the increased anaerobic oxidation of glucose towards lactic acid, it seems unlikely that sorbitol accumulation was due to the activation of sorbitol synthesis. (2) The next step in the polyol pathway is the degradation of sorbitol to fructose, a reaction catalysed by the NAD^+ -dependent enzyme sorbitol dehydrogenase (SD). In this study, sorbitol accumulation may be attributed to the alcohol-induced inhibition of its catabolism — due to the metabolism of

ethanol, the NAD^+ required for the dehydrogenation of sorbitol becomes depleted. Furthermore, significant down-regulation of SD has been observed in a study on the changes of the cytoplasmic proteome in response to alcoholic hepatotoxicity in rats [44], which lends support to this viewpoint. (3) Increased sorbitol as a consequence of its role in osmoregulation, however, seems to be the preferred explanation for interpreting the observed urinary excretion profile of sorbitol. Ethanol is both water and lipid soluble, which renders it a membrane-permeable substance. Abnormal cell volume regulation significantly contributes to the pathophysiology of several disorders, and cells respond to these changes by importing, exporting, or synthesizing osmolytes to maintain volume homeostasis [45]. On an experimental level, oedema/cell swelling could be induced by binge-simulated ethanol exposure in slice cultures of the developing rat brain [46]. In a cell-to-medium flux study, hyperosmotically induced intracellular accumulation of sorbitol in renal epithelial cells showed a greater than 150-fold increased efflux within five minutes after exposure to an isosmotic medium. We thus speculate that the increased sorbitol excretion in our alcohol-exposed subjects relates to ethanol-induced hepatocyte swelling, which is compensated for by sorbitol release in order to maintain cellular volume homeostasis. Although less pronounced in this study than that of sorbitol, perturbations of other osmolytes, such as TMAO and taurine (observed in this study, see Table 2), should also be considered in the examination of osmolyte responses.

In summary, we have described the urinary metabolite profile of healthy, young males following acute alcohol consumption as part of a designed intervention study. The complex NMR spectral data, generated from individuals participating in a time-dependent cross-over study, could be resolved sufficiently through the application of univariate and multivariate statistical analyses. This approach provided a novel method for the analysis and understanding of the complex metabolomics data that were produced due to acute alcohol consumption. We indicate that NAD^+ depletion, and the production of an excessive amount of reducing equivalents, greatly perturb the hepatocyte redox homeostasis, resulting in metabolic disturbances, of which urinary excretion of sorbitol is novel. We postulate that sorbitol is a marker of a cell volume regulatory response to ethanol-induced hepatocyte swelling. This may have a wider significance related to brain oedema-induced neurodegenerative damage following chronic binge alcohol exposure [46,47].

Acknowledgements

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 TIA. The funder was not involved in the writing of the manuscript, neither in its editing nor in the decision for it to be distributed for publication.

Competing interests

The authors declare that there is no conflict of interest.

Author Contributions

Conceptualization: CJR, JAW. **Data curation:** CI, SM, MvR, JAW. **Formal analysis:** CI, SM. **Funding acquisition:** CJR. **Investigation:** CI, SM. **Methodology:** CJR, JAW. **Project administration:** CJR, JAW. **Software:** JAW, MvR. **Supervision:** CJR, JAW. **Writing & original draft:** CI, MvR, SM, LJM, RW, JAW, CJR. **Writing & review & editing:** CI, MvR, SM, LJM, RW, JAW, CJR.

Supporting information

S1 File. Details of sample preparation, ¹H-NMR analysis and data analysis methods.

Section A: Method for sample treatment, storage, preparation and ¹H-NMR analysis

Section A.1: Sample collection, characterization and storage

Section A.2: Sample preparation and ¹H-NMR analysis

Section A.3: Data processing

Section A.3.1 Pre-processing

Section A.3.2 Quality assurance

Section A.3.3 The interaction effect of NAD and alcohol

Section B: Uric acid analysis

Section C: Original ¹H-NMR spectral data

References

S2 File. Raw ¹H-NMR spectral data are given as an electronic file in Excel format (S2_File.xlsx).

References

1. Gibney MJ, Walsh M, Brennan L, Roche HM, German B, Van Ommen B. Metabolomics in human nutrition: opportunities and challenges. *Am J Clin Nutr.* 2005; 82:497–503.
2. Wishart DS. Metabolomics: applications to food science and nutrition research. *Trends Food Sci Tech.* 2008; 19:482–493.
3. Irwin C, van Reenen M, Mason S, Mienie LJ, Westerhuis JA, Reinecke CJ. Contribution towards a metabolite profile of the detoxification of benzoic acid through glycine conjugation: an intervention study. *PLOS ONE.* 2016; 11(12):e0167309. doi:10.1371/journal.pone.0167309
4. Bradford BU, Rusyn I. Swift increase in alcohol metabolism (SIAM): understanding the phenomenon of hypermetabolism in liver. *Alcohol.* 2005; 35(1):13–17.

5. Hawkins RD, Kalant H. The metabolism of ethanol and its metabolic effects. *Pharmacol Rev.* 1972; 24(1):67–157.
6. Zakhari S, Li T-K. Determinants of alcohol use and abuse: impact of quantity and frequency patterns on liver disease. *Hepatology.* 2007; 46(6):2032–2039.
7. Lieber CS. Ethanol metabolism, cirrhosis and alcoholism. *Clin chim acta.* 1997; 257(1):59–84.
8. Nahon P, Amathieu R, Triba MN, Bouchemal N, Nault JC, Ziol M, et al. Identification of serum proton NMR metabolomic fingerprints associated with hepatocellular carcinoma in patients with alcoholic cirrhosis. *Clin Cancer Res.* 2012; 18(24):6714–6722.
9. Rachakonda V, Gabbert C, Raina A, Bell LN, Cooper S, Malik S, et al. Serum metabolomic profiling in acute alcoholic hepatitis identifies multiple dysregulated pathways. *PLOS ONE.* 2014; 9(12):e113860
10. Amathieu R, Triba MN, Nahon P, Bouchemal N, Kamoun W, Haouache H, et al. Serum ¹H-NMR metabolomic fingerprints of acute-on-chronic liver failure in intensive care unit patients with alcoholic cirrhosis. *PLOS ONE.* 2014; 9(2):e89230.
11. Lian JS, Liu W, Hao SR, Guo YZ, Huang HJ, Chen DY, et al. A serum metabonomic study on the difference between alcohol- and HBV-induced liver cirrhosis by ultraperformance liquid chromatography coupled to mass spectrometry plus quadrupole time-of-flight mass spectrometry. *Chin Med J.* 2011; 124(9):1367–1373.
12. Gika HG, Ji C, Theodoridis GA, Michopoulos F, Kaplowitz N, Wilson ID. Investigation of chronic alcohol consumption in rodents via ultra-high-performance liquid chromatography–mass spectrometry based metabolite profiling. *J Chromatogr A.* 2012; 1259:128–137.
13. Loftus N, Barnes A, Ashton S, Michopoulos F, Theodoridis G, Wilson I, et al. Metabonomic investigation of liver profiles of nonpolar metabolites obtained from alcohol-dosed rats and mice using high mass accuracy MSn analysis. *J Proteome Res.* 2010; 10(2):705–713.
14. Gao X, Zhao A, Zhou M, Lin J, Qiu Y, Su M, et al. GC/MS-based urinary metabolomics reveals systematic differences in metabolism and ethanol response between Sprague–Dawley and Wistar rats. *Metabolomics.* 2011; 7(3) 363–374.
15. Kirpich IA, Petrosino J, Ajami N, Feng W, Wang Y, Liu Y, et al. Saturated and unsaturated dietary fats differentially modulate ethanol-induced changes in gut microbiome and metabolome in a mouse model of alcoholic liver disease. *Am J Pathol.* 2016; 186:765–776.

16. Manna SK, Patterson AD, Yang Q, Krausz KW, Idle JR, Fornace Jr. AJ, et al. UPLC–MS-based urine metabolomics reveals indole-3-lactic acid and phenyllactic acid as conserved biomarkers for alcohol-induced liver disease in the *Ppara*-null mouse model. *J Proteome Res.* 2011; 10(9), 4120–4133.
17. Gika HG, Wilson ID. Metabolic profiling approaches for biomarkers of ethanol intake. In: Patel VB, editor. *Molecular aspects of alcohol and nutrition.* Academic Press; 2016. pp. 213–222.
18. Nicholas PC, Kim D, Crews FT, Macdonald JM. ¹H NMR-based metabolomic analysis of liver, serum, and brain following ethanol administration in rats. *Chem Res Toxicol.* 2008; 21:408–420.
19. Bradford BU, O'Connell TM, Han J, Kosyk O, Shymonyak S, Ross PK, et al. Metabolomic profiling of a modified alcohol liquid diet model for liver injury in the mouse uncovers new markers of disease. *Toxicol Appl Pharmacol.* 2008; 232:236–243.
20. Masuo Y, Imai T, Shibato J, Hirano M, Jones OAH, Maguire ML, et al. Omic analyses unravels global molecular changes in the brain and liver of a rat model for chronic Sake (Japanese alcoholic beverage) intake. *Electrophoresis.* 2009; 30:1259–1275.
21. Fernando H, Kondraganti S, Bhopale KK, Volk DE, Neerathilingam M, Kaphalia BS, et al. ¹H and ³¹P NMR lipidome of ethanol-induced fatty liver. *Alcohol Clin Exp Res.* 2010; 34(11):1937–1947.
22. Yoseph BP, Breed E, Overgaard CE, Ward CJ, Liang Z, Wagener ME, et al. Chronic alcohol ingestion increases mortality and organ injury in a murine model of septic peritonitis. *PLOS ONE.* 2013; 8(5):e62792. doi:10.1371/journal.pone.0062792
23. Vázquez-Fresno R, Llorach R, Alcaro F, Rodríguez MA, Vinaixa M, Chiva-Blanch G, et al. (1)H-NMR-based metabolomic analysis of the effect of moderate wine consumption on subjects with cardiovascular risk factors. *Electrophoresis.* 2012; 33(15):2345–2354.
24. Würtz P, Cook S, Wang Q, Tiainen M, Tynkkynen T, Kangas AJ, et al. Metabolic profiling of alcohol consumption in 9778 young adults. *Int J Epidemiol.* 2016; 45(5):1493–1506.
25. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, et al. The human urine metabolome. *PLOS ONE.* 2013; 8(9):e73076. doi: 10.1371/journal.pone.0073076.
26. Bala S, Marcos M, Gattu A, Catalano D, Szabo G. Acute binge drinking increases serum endotoxin and bacterial DNA levels in healthy individuals. *PLOS ONE.* 2014; 9(5): e96864. doi:10.1371/journal.pone.0096864
27. Antti H, Bollard ME, Ebbels T, Keun H, Lindon JC, Nicholson JK, et al. Batch statistical processing of 1NMR-derived urinary spectral data. *J Chemometrics.* 2002; 16:461–468.

28. Smilde AK, Jansen JJ, Hoefsloot HCJ, Lamers R-JAN, Van der Greef J, Timmerman ME. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics*. 2005; 21:3043–3048.
29. World Health Organization. Global status report on alcohol and health 2014. http://www.who.int/substance_abuse/publications/global_alcohol_report/en/ Date of access: 24 Aug. 2017.
30. Numminen H, Syrjälä M, Benthin G, Kaste M, Hillborn M. The effect of acute ingestion of a large dose of alcohol on the hemostatic system and its circadian variation. *Stroke*. 2000, 31(6):1269–1273.
31. Viant M, Bearden DW, Bundy JG, Burton IW, Colette TW, Ekman DR, et al. International NMR-based environmental metabolomics intercomparison exercise. *Environ Sci Technol*. 2009; 43:219–225.
32. Westerhuis JA, van Velzen EJJ, Hoefsloot HCJ, Smilde AK. Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. *Metabolomics*. 2010; 6(1):119–128.
33. Lieber CS. Alcohol and the liver: 1994 update. *Gastroenterology*. 1994; 106(4):1085–1105.
34. Day CP, Bashir R, James OF, Bassendine MF, Crabb DW, Thomasson HR, et al. Investigation of the role of polymorphisms at the alcohol and aldehyde dehydrogenase loci in genetic predisposition to alcohol-related end-organ damage. *Hepatology*. 1991; 14(5):798–801.
35. Yamamoto T, Moriwaki Y, Takahashi S. Effect of ethanol on metabolism of purine bases (hypoxanthine, xanthine, and uric acid). *Clin Chim Acta*. 2005; 356(1):35–57.
36. Yuki T, Thurman RG. The swift increase in alcohol metabolism: Time course for the increase in hepatic oxygen uptake and the involvement of glycolysis. *Biochem J*. 1980; 186(1):119–126.
37. Faller J, Fox IH. Ethanol-induced hyperuricemia: evidence for increased urate production by activation of adenine nucleotide turnover. *N Engl J Med*. 1982; 307(26):1598–1602.
38. Puig JG, Fox IH. Ethanol-induced activation of adenine nucleotide turnover: Evidence for the role of acetate. *J Clin Invest*. 1984; 74(3):936–941.
39. Nishimura T, Shimizu T, Mineo I, Kawachi M, Ono A, Nakajima H, et al. Influence of daily drinking habits on ethanol-induced hyperuricemia. *Metabolism*. 1994; 43(6):745–748.
40. Yü TF, Sirota JH, Berger L, Halpern M, Gutman AB. Effect of Sodium Lactate Infusion on Urate Clearance in Man. *Proc Soc Exp Biol Med*. 1957; 96(3):809–813.
41. Strange K. Cellular volume homeostasis. *Adv Physiol Educ*. 2004; 28:155–159. doi:10.1152/advan.00034.2004

42. Lang F, Busch GL, Ritter M, Völkl H, Waldegger S, Gulbins E, et al. Functional significance of cell volume regulatory mechanisms. *Physiol rev.* 1998; 78(1):247–306.
43. Tang WH, Martin KA, Hwa J. Aldose reductase, oxidative stress, and diabetic mellitus. *Front Pharmacol.* 2012; 3:87. doi: 10.3389/fphar.2012.00087
44. Kim DH, Lee E-M, Do S-H, Jeong D-H, Jeong KS. Changes of the Cytoplasmic Proteome in Response to Alcoholic Hepatotoxicity in Rats. *Int J Mol Sci.* 2015; 16:18664–18682. doi:10.3390/ijms160818664
45. Diamond I, Gordon AS. Cellular and molecular neuroscience of alcoholism. *Physiol rev.* 1997; 77(1):1–20.
46. Moon KH, Tajuddin N, Brown J, Neafsey EJ, Kim HY, Collins MA. Phospholipase A2, Oxidative Stress, and Neurodegeneration in Binge Ethanol-Treated Organotypic Slice Cultures of Developing Rat Brain. *Alcohol Clin Exp Res.* 2014; 38(1):161–169.
47. Collins MA, Zou JY, Neafsey EJ. Brain damage due to episodic alcohol exposure in vivo and in vitro: furosemide neuroprotection implicates edema-based mechanism. *FASEB J.* 1998; 12:221–230.

The ¹H-NMR-based metabolite profile of acute alcohol consumption: a metabolomics intervention study — Supporting Information

Cindy Irwin^{1¶}, Mari van Reenen^{1,2¶}, Shayne Mason¹, Lodewyk J. Mienie¹, Ron A. Wevers³, Johan A. Westerhuis^{2,4} and Carolus J. Reinecke^{1*}

¹Centre for Human Metabolomics, Faculty of Natural Sciences and Agriculture, North-West University (Potchefstroom Campus), Potchefstroom, South Africa

²Department of Statistics, Faculty of Natural Sciences and Agriculture, North-West University (Potchefstroom Campus), Potchefstroom, South Africa

³Department of Laboratory Medicine, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

⁴Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands

*Corresponding author

E-mail: carools.reinecke@nwu.ac.za (CJ)

¶These authors contributed equally to this work

Section A Methods for sample treatment, storage, preparation and ^1H -NMR analysis

Section A.1 Sample collection, characterization and storage

All the subjects were randomly assigned to one of the four intervention groups until all 24 had participated in all four interventions. Six urine samples were collected from each subject for each intervention, giving a total of 24 samples from each subject over the course of the study.

One 5 mL and two 1 mL vials were used to provide aliquots of each of the urine samples; these aliquots, together with the remainder of the bulk urine samples, were stored at -80°C . Once all the urine samples were collected, one 1 mL aliquot of each was thawed and combined to prepare a pooled quality control (QC) sample for the experiment as a whole. This QC sample was then divided into 15 mL aliquots and once again stored at -80°C .

The collected samples were analysed in 24 separate batches, each batch containing the 24 samples of a single subject and three QC samples. The 27 samples from each of the 24 batches were analysed in the following order:

$\text{QC}_1 [\text{S}_{-1}\text{S}_0\text{S}_1\text{S}_2\text{S}_3\text{S}_4]_{\text{Vehicle}} [\text{S}_{-1}\text{S}_0\text{S}_1\text{S}_2\text{S}_3\text{S}_4]_{\text{Alcohol}} \text{QC}_2 [\text{S}_{-1}\text{S}_0\text{S}_1\text{S}_2\text{S}_3\text{S}_4]_{\text{NAD}}$

$[\text{S}_{-1}\text{S}_0\text{S}_1\text{S}_2\text{S}_3\text{S}_4]_{\text{NAD+Alcohol}} \text{QC}_3$

where S_{-1} represents the sample collected at time -1 , S_0 represents the sample collected at time 0, and so on.

Section A.2 Sample preparation and ^1H -NMR analysis

Spectral analyses were conducted at the NMR facility of the Centre for Human Metabolomics at North-West University. Prior to analysis, an aqueous 1.5 M KH_2PO_4 deuterated buffer solution at pH 7.4 was prepared [1]. This solution served to lock the signal during analysis, maintained a stable pH in the sample and contained trimethyl-2,2,3,3-tetradeuteriopropionic acid (TSP, sodium salt; Sigma Aldrich) as the internal standard to provide a chemical shift reference of $\delta = 0.00$. The urine samples, stored at -80°C , were thawed at room temperature for analysis. A 600 μL volume of each sample was centrifuged at $12\,000 \times g$ for 5 min to remove any sediments or debris. A 60 μL volume of buffer solution was added to 540 μL of the supernatant, vortexed and transferred to a 5-mm NMR tube.

Each sample so prepared was analysed on a Bruker Avance III HD 500 MHz NMR spectrometer equipped with a triple-resonance inverse (TXI) $^1\text{H}\{^{15}\text{N}, ^{13}\text{C}\}$ probe head and x, y, z gradient coils. ^1H spectra were acquired as 128 transients in 32K data points with a spectral width of 6002 Hz. The sample temperature was maintained at 300 K and the H_2O

resonance was pre-saturated by single-frequency irradiation during a relaxation delay of 4 s, with a 90° excitation pulse fixed at 8 μ s. Shimming of the sample was performed automatically on the deuterium signal. The resonance line widths for TSP and metabolites were <1 Hz (measurements at half the height of the peak). Fourier transformation and phase and baseline correction were done automatically. The software used was Bruker Topspin (V3.2) and Bruker AMIX (V3.9.9) [2].

All urine samples were normalized with reference to the creatinine CH₂ peak at 4.05 ppm. We employed two methods of spectral analysis: (1) Applying equidistant binning (using a bin width of 0.02 ppm) to the spectral region between 0.5 and 10 ppm, excluding the region of the water peak (4.72–4.88 ppm). This gave a total of 467 integrated units per NMR spectrum for each sample for statistical analysis — spectral data. (2) Using information from the statistical outputs to identify discriminatory metabolites (based upon pure compound spectral libraries), and then accurately integrating the selected metabolites — providing quantified concentration data (μ mol metabolite/mmol creatinine) for univariate analysis (see Table 2 in the main text).

Section A.3 Data processing

A.3.1 Pre-processing

The use of NMR spectroscopy as a tool for metabolomics is limited by the sensitivity (limit of detection) of NMR [3], requiring an approach in discarding noise while retaining and then measuring real peaks. The spectral intensities within the original 467 bins generated here ranged from 7.8×10^5 to 1.7×10^9 . Based on previous empirical experience with NMR spectral analysis, we defined a threshold value of 2×10^6 , being approximately the limit of detection of metabolomic substances presumed to be present in a spectral bin. To reduce the uncertainty in the data at the level of the approximate detection limit, we applied the threshold of 2×10^6 to the data matrix, setting all values below this threshold to zero. We subsequently applied a “zero-filter” to remove bins with more than 50% zero observations across all six time points, as applied for fusion of MS-based metabolomics data [4]. This resulted in a reduction in the number of bins containing spectral data from 467 to 347.

The remaining zero observations were replaced by random numbers below the minimum non-zero observation for each bin, after which the data were log scaled and auto scaled.

To account for the dilution differences between the urine samples, the value of each bin was normalized relative to the CH₂ peak of creatinine. This peak is contained in the 4.05 and 4.07 ppm bins. Therefore, each bin value was divided by the sum of the value in these two bins to normalize the bin values relative to creatinine. After normalization, the three bins related to creatinine (4.05, 4.07 and 3.05 ppm) were also removed from the matrix, giving a final total of 344 bins containing spectral data.

A.3.2 Quality assurance

Quality control (QC) samples were included in each batch at the start, middle and end of the analytical run. The variation and correlation within and between bins for QC and experimental samples were compared using principal component analysis (PCA) to attain a birds-eye view of any batch effects. NMR data are not known typically to present with such variation structures (that is, within or between batch effects), and this was also the case here. The PCA scores plot (Figure A) shows a similar correlation structure for the QC and experimental samples, but much less variation in the QC samples. From these results it was assumed that no batch corrections were required.

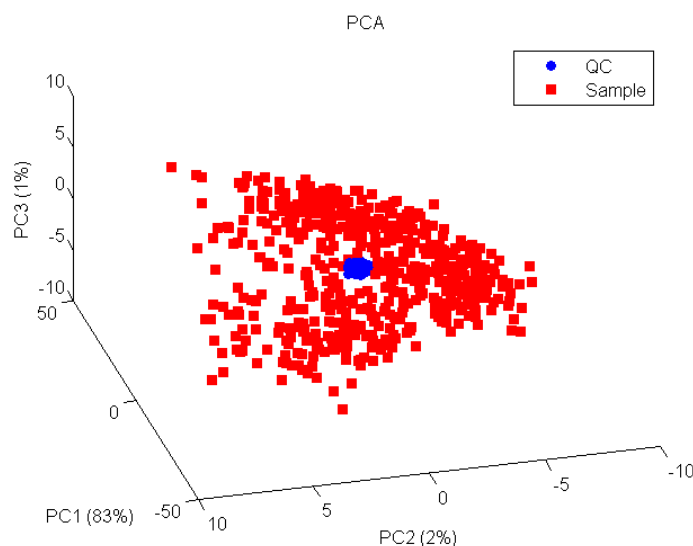


Figure A. PCA scores plot illustrating the variation and correlation within and between bins for QC and experimental samples. The close clustering of the QC samples compared to the experimental samples indicates that the variation in the QC samples is much less than the variation in the experimental samples. These results suggest that no batch corrections were required.

A.3.3 The interaction effect of NAD and alcohol

As explained in the main text, the complete design of the study (that is, four intervention measured over 6 time points across all individuals) was modelled for quantified hypoxanthine and sorbitol using a two-way repeated measures analysis of variance (2-way RM ANOVA) model. A brief overview of the information extracted from the model is presented in Fig 5 of the main text. The absence of these metabolites prior to the intervention and sudden increase in their levels shortly thereafter, resulted in a high frequency of zero observations, thus introducing a spike in the distribution of the data. Therefore, the nature of the experiment and metabolites under investigation resulted in the data not being normally distributed. To account for this, log transformed data were used to build the RM ANOVA model, while Greenhouse–Geisser-corrected p-values were used to assess the significance of the main effects (that is, intervention and time). Unfortunately,

inspection of the residuals indicated that the transformation was not able to produce normality across all effect levels. The Wilcoxon signed-rank test was used as a confirmation test on the untransformed data to test for significantly different pairs of means as indicated in Figs 5a and 5c in the main text. However, since the performance of the Wilcoxon signed-rank test may also be compromised due to the spike in the data, difference data were assessed as a final confirmation to provide additional proof of the differences observed. The difference data were created by subtracting time 0 data from the remaining time points, while ignoring time -1 . Doing so significantly improved the distribution of the data. The results based on the difference data are summarised in Figs 5b and 5d in the main text.

Section B Uric acid analysis

The manual of the protocol detailing the uric acid analysis is available online as a downloadable PDF at <https://www.thermofisher.com/order/catalog/product/TR24321> [5].

Infinity™ Uric Acid Liquid Stable Reagent

PRODUCT SUMMARY

Stability	:	Until Expiry at 2 - 8°C
Linear Range	:	0.03-1.5 mmol/L (0.5-25.2 mg/dL)
Specimen Type	:	Serum or Urine
Method	:	Enzymatic Endpoint
Reagent Preparation	:	Supplied ready to use.

IVD

INTENDED USE

This reagent is intended for the in vitro quantitative determination of Uric Acid in human serum or urine.

CLINICAL SIGNIFICANCE

Uric acid is a metabolite of purines, nucleic acids and nucleoproteins: consequently, abnormal levels may be indicative of a disorder in the metabolism of these substances. Hyperuricaemia may be observed in renal dysfunction, gout, leukemia, polycythaemia, atherosclerosis, diabetes, hypothyroidism, or in some genetic diseases. Decreased levels are present in patients with Wilson's Disease.^{1, 2, 3}

METHODOLOGY

This reagent is based upon the methods of Trivedi and Kabasakalian^{4, 5} with a modified Trinder⁶ peroxide assay using 2,4,6-Tribromo-3-hydroxy benzoic acid (TBHB).

The series of reactions involved in the assay system is as follows:

1. $\text{Uric Acid} + \text{O}_2 + \text{H}_2\text{O} \xrightarrow{\text{Uricase}} \text{Allantoin} + \text{CO}_2 + \text{H}_2\text{O}_2$

2. $2\text{H}_2\text{O}_2 + 4\text{-AAP} + \text{TBHB} \xrightarrow{\text{Peroxidase}} \text{Quinoneimine} + \text{H}_2\text{O}$

1. Uric Acid is oxidised to allantoin by uricase with the production of H₂O₂.
2. The peroxide reacts with 4-aminoantipyrine (4-AAP) and TBHB in the presence of peroxidase to yield a quinoneimine dye. The resulting change in absorbance at 520nm (520-550nm) is proportional to uric acid concentration in the sample.

REAGENT COMPOSITION

Active Ingredients

	Concentration
4-Aminoantipyrine	0.5 mmol/L
TBHB	1.75 mmol/L
Uricase (Bacillus Sp.)	> 120 U/L
Peroxidase (Horseradish)	> 500 U/L
Tris Buffer	50 mmol/L

pH 8.25 ± 0.1 at 20°C.

WARNING: Do not ingest. Avoid contact with skin and eyes. If spilt thoroughly wash affected areas with water. Reagent contains Sodium Azide which may react with copper or lead plumbing. Flush with plenty of water when disposing. For further information consult the Infinity Uric Acid Liquid Stable Reagent Material Safety Data Sheet.

CAUTION: This product contains animal source material. Handle and dispose of this product as if it were potentially infectious.

REAGENT PREPARATION

Reagent is supplied ready to use.

STABILITY AND STORAGE

When stored refrigerated at 2-8°C the reagent is stable until the expiry date stated on the bottle and kit box label.

Indications of Reagent Deterioration:

- Turbidity;
- Reagent Absorbance >0.5 AU at 520nm; and/or
- Failure to recover control values within the assigned range.

SPECIMEN COLLECTION AND HANDLING

Collection: No special preparation of the patient is required.

SYMBOLS IN PRODUCT LABELLING

EC	Authorized Representative		Temperature Limitation
IVD	For in vitro diagnostic use		Use by/Expiration Date
LOT	Batch code/Lot number		CAUTION. CONSULT INSTRUCTIONS FOR USE.
REF	Catalogue number		Manufactured by
	Consult instructions for use		

Serum: Use non-haemolysed serum.

Urine: It is recommended that 15 mL of 2 mol/L NaOH be added to the collection vessel. Upon receipt of the urine sample, pH should be checked. If the pH is less than 8.0 it should be adjusted with NaOH. A 1:10 dilution of urine is typically required prior to analysis.⁷

Storage: Serum samples are stable for at least 3 days at room temperature (18-25°C) and for at least 6 months frozen.² Stabilized urine may be stored at room temperature for 5 days.⁷

ADDITIONAL EQUIPMENT REQUIRED BUT NOT PROVIDED

- If required, pipettes for accurately dispensing measured volumes.
- A clinical chemistry analyzer capable of maintaining constant temperature (37°C) and measuring absorbance at 520 nm.
- Analyzer specific consumables, eg: sample cups.
- Normal and abnormal assayed control material.
- Calibrator or a suitable aqueous standard (see calibration section).

ASSAY PROCEDURE

The following system parameters are recommended. Individual instrument applications are available upon request from the Technical Support Group.

SYSTEM PARAMETERS

Temperature	37°C
Primary Wavelength	520 nm (520-550 nm)
Secondary Wavelength	600 - 660 nm
Assay Type	Endpoint
Direction	Increase
Sample : Reagent Ratio	1 : 50
eg: Sample Vol	3 µL
Reagent Vol	150 µL
Incubation Time	300 Seconds
Reagent Blank Limits	Low 0.0 AU
(520nm, 1cm light path)	High 0.5 AU
Linearity	0.03-1.50 mmol/L (0.5-25.2 mg/dL)
Analytical Sensitivity	0.42 ΔA per mmol/L
(520nm, 1cm light path)	0.025ΔA per mg/dL

CALCULATIONS

Results are calculated, usually automatically by the instrument, as follows:

$$\text{Uric Acid} = \frac{\text{Absorbance of Unknown}}{\text{Absorbance of Calibrator}} \times \text{Calibrator Value}$$

Example:

Absorbance of calibrator	= 0.302
Absorbance of unknown	= 0.071
Value of calibrator	= 0.720 mmol/L (12.1 mg/dL)

$$\text{Uric Acid} = \frac{0.071}{0.302} \times 0.720 = 0.16 \text{ mmol/L}$$

$$\text{Uric Acid} = \frac{0.071}{0.302} \times 12.1 = 2.8 \text{ mg/dL}$$

NOTES

1. The reagent and sample volumes may be altered proportionally to accommodate different spectrophotometer requirements.
2. The color development is stable for 15 minutes.
3. Specimens with Uric Acid concentrations greater than 1.50 mmol/L (25.2 mg/dL) should be diluted with saline and reassayed. Multiply results by the dilution factor.

Thermo
SCIENTIFIC

4. S.I. unit conversion factor: mmol/L x 16.8 = mg/dL.

CALIBRATION

Calibration is required. An aqueous standard or serum based calibrator, with an assigned value traceable to a primary standard (eg NIST or IRMM) is recommended. Standards should not contain formaldehyde or enzyme inhibitors as preservatives. For calibration frequency on automated instruments, refer to the instrument manufacturers specifications.

However, calibration stability is contingent upon optimum instrument performance and the use of reagents which have been stored as recommended in the stability and storage section of this package insert. Recalibration is recommended at anytime if one of the following events occurs:-

- The lot number of reagent changes.
- Preventative maintenance is performed or a critical component is replaced.
- Control values have shifted or are out of range and a new vial of control does not rectify the problem.

QUALITY CONTROL

To ensure adequate quality control, normal and abnormal control with assayed values should be run as unknown samples:-

- At least once per day or as established by the laboratory.
- When a new bottle of reagent is used.
- After preventative maintenance is performed or a critical component is replaced.
- With every calibration.

Control results falling outside the upper or lower limits of the established ranges indicate the assay may be out of control.

The following corrective actions are recommended in such situations:-

- Repeat the same controls.
- If repeated control results are outside the limits, prepare fresh control serum and repeat the test.
- If results are still out of control, recalibrate with fresh calibrator, then repeat the test.
- If results are still out of control, perform a calibration with freshly prepared reagent, then repeat the test.
- If results are still out of control, contact Technical Services or the local distributor.

LIMITATIONS

1. Studies to determine the level of interference from haemoglobin, bilirubin (free and conjugated) and lipaemia were carried out. The following results were obtained:

Haemoglobin: No interference from haemoglobin up to 424 mg/dL.

Free Bilirubin: No interference from free bilirubin up to 212 µmol/L (12 mg/dL).

Conjugated Bilirubin: No interference from conjugated bilirubin up to 212 µmol/L (12 mg/dL).

Lipaemia: No interference from lipaemia, measured as absorbance at 630nm, up to 1.68 AU.

2. Young DS² has published a comprehensive list of drugs and substances which may interfere with this assay.

EXPECTED VALUES *

Child:	0.12 - 0.32 mmol/L	2.0 - 5.0 mg/dL
Adult Male:	0.21 - 0.42 mmol/L	3.5 - 7.2 mg/dL
Adult Female:	0.15 - 0.35 mmol/L	2.6 - 6.0 mg/dL
Urine:	1.48 - 4.43 mmol/day	250 - 750 mg/day

The quoted values should serve as a guide only. It is recommended that each Laboratory verify this range or derives a reference interval for the population it serves.¹⁰

PERFORMANCE DATA

The following data was obtained using the Infinity Uric Acid Liquid Stable Reagent on a well maintained automated clinical chemistry analyzer. Users should establish product performance on their specific analyzer used.

IMPRECISION

Imprecision was evaluated over a period of 20 days using two levels of commercial

control and following the NCCLS EP5-T procedure.¹¹

	LEVEL I	LEVEL II
Number of data points	80	80
Mean (mmol/L / mg/dL)	0.279 / 4.69	0.603 / 10.13
Within run: SD (mmol/L / mg/dL)	0.007 / 0.12	0.009 / 0.15
CV (%)	2.3	1.5
Total: SD (mmol/L / mg/dL)	0.019 / 0.32	0.021 / 0.35
CV (%)	6.8	3.4

METHOD COMPARISON

Comparison studies were carried out using a similar commercially available reagent as a reference. Serum and urine samples were assayed in parallel and the results compared by least squares regression. The following statistics were obtained.

Serum:

Number of sample pairs	60
Range of sample results	0.11-0.61 mmol/L (1.85-10.25 mg/dL)
Mean of reference method results	0.315 mmol/L (5.29 mg/dL)
Mean of Uric Acid results	0.336 mmol/L (5.65 mg/dL)
Slope	0.931
Intercept	0.042 mmol/L (0.71 mg/dL)
Correlation coefficient	0.987

Urine:

Number of sample pairs	41
Range of sample results	0.48 - 11.7 mmol/L (8.0 - 196 mg/dL)
Mean of reference method results	3.0 mmol/L (49.6 mg/dL)
Mean of Uric Acid results	2.5 mmol/L (42.7 mg/dL)
Slope	0.967
Intercept	-0.32 mmol/L (-5.3 mg/dL)
Correlation coefficient	0.990

LINEARITY

When run as recommended the assay is linear between 0.03 and 1.50 mmol/L (0.5-25.2 mg/dL).

ANALYTICAL SENSITIVITY

When run as recommended the sensitivity of this assay is 0.42 ΔAbs per mmol/L or 0.025 ΔAbs per mg/dL (1cm light path, 520nm).

REFERENCES

1. Searcy R.L., Diagnostic Biochemistry. McGraw-Hill, New York, NY, 1969.
2. Henry R.J., Common C. and Winkelman J.W. (eds), Clinical Chemistry: Principles and Techniques. Harper & Row, Hagerstown, MD, 1974.
3. Balis M. E., Adv. Clin. Chem. 18(213) 1976.
4. Trivedi R., Berta E. and Rebar L., Clin. Chem. 22(1223), 1976.
5. Kabasakalian P. Kalliney S. and Wescott A. Clin. Chem. 19(522) 1973.
6. Trinder P., J. Clin. Pathol. 22(246) 1949.
7. Shephard MD, Mezzachhi RD. Clin Biochem Revs 1983; 4:61-7.
8. Young DS. Effects of Drugs on Clinical Laboratory Tests. Third Edition. 1990; 3:360-370.
9. Tietz Textbook of Clinical Chemistry and Molecular Diagnosis (4th Ed.) Burtis, Ashwood & Bruns (Eds), Elsevier Saunders, 2005; 2301.
10. Wachtel M et al. Creation and Verification of Reference Intervals. Laboratory Medicine 1995; 26:593-7.
11. National Committee for Clinical Laboratory Standards. User evaluation of Precision Performance of Clinical Chemistry Devices. NCCLS; 1984, NCCLS Publication EP5-T.



Fisher Diagnostics
a division of Fisher Scientific Company, LLC
a part of Thermo Fisher Scientific Inc.
Middletown, VA 22645-1905 USA
Phone: 800-528-0494
540-869-3200
Fax: 540-869-8132



WMDE
Bergerweg 18
6085 AT Horn
The Netherlands

JL840768-en (R1)



© 2012 Thermo Fisher Scientific Inc. All rights reserved. Hitachi is a registered trademark of Roche Diagnostics, Indianapolis, IN 46250. iLab 600 is a registered trademark of Instrumentation Laboratory Company, Lexington, MA 02421. All other trademarks are the property of Thermo Fisher Scientific Inc and its subsidiaries.



Reorder Information

Catalogue No.	Configuration
TR24321	2 x 125 mL

Section C Original ¹H-NMR spectral data

The data include urine samples from all 24 experimental subjects and all four interventions (576 samples), as well as 72 QC samples, yielding a total of 648 study samples. Data from four of the subjects were, however, removed, as explained in the main text. The spectral region between 0.5 and 10 ppm, excluding the region of the water peak (4.72–4.88 ppm), of the original ¹H-NMR spectrum of each of the urine samples analysed was divided into 0.02-ppm equal-sized bins. This resulted in 467 bins containing spectral data. The original, raw spectral data matrix of all the samples analysed for this study are given as an electronic file in Excel format (S2_File.xlsx). Table A represents a small extract from this file.

Table A. A small extract from the file containing the raw ¹H-NMR spectral data.

Batch	Sample Name	Case	Treatment	Time	Bin: 1	2	3	...	466	467
					Bin no: 9.99	9.969999	9.95	...	0.53	0.51
1	QC1	F	QC		144185.2	85175.12	97222.12		162878.1	164512.9
1	F41-1	F	Vehicle	–1	252890.2	171823.4	163215.5		384216.4	266010.8
1	F410	F	Vehicle	0	151299.1	100954.3	86908.15		262467	216768.2
1	F411	F	Vehicle	1	79268.87	105579.2	95048.77		76325.53	126937.3
1	F412	F	Vehicle	2	84927.78	101751.3	127028.2		77872.51	68765.72
1	F413	F	Vehicle	3	74747.86	144056.7	47937.61		91158.41	58079.11
1	F414	F	Vehicle	4	67488.71	70024.98	68237.62		103052	89226.37
1	F32-1	F	Alcohol	–1	140557.8	103222.9	98290.31		297911.4	217084.5
1	F320	F	Alcohol	0	186432.5	118033	86594.67		224899.6	205953.7
1	F321	F	Alcohol	1	138804.5	89211.63	82459.67		137392	87124.63
1	F322	F	Alcohol	2	74946.59	76699.26	70328.33		125527.2	139514.5
1	F323	F	Alcohol	3	70183.16	75978.23	71975.92		90170.89	86251.59
1	F324	F	Alcohol	4	170387.9	119688.5	133812.7		125650	97010.52
1	QC2	F	QC		148854.6	107287.3	109676.4		194299.3	116989.8
1	F22-1	F	NAD	–1	86274.71	155479	145948.6		176155.9	165445.7
1	F220	F	NAD	0	137788.9	91139.97	80084.29		169180.1	168179.4
1	F221	F	NAD	1	101979	64201.66	84790.44		92116.46	112283.7
1	F222	F	NAD	2	81776.94	62372.04	73383.53		94053.74	71139.18
1	F223	F	NAD	3	103830	99023.64	83557.65		87499.72	105244.2
1	F224	F	NAD	4	81614.46	71494.57	85544.25		99941.91	85364.31
1	F14-1	F	NAD + Alcohol	–1	167690.9	147427.5	99771.65		171520.8	176960.6
1	F140	F	NAD + Alcohol	0	79105.6	75316.07	71888.47		111565.1	74134.41
1	F141	F	NAD + Alcohol	1	67927.37	97794.58	50990.82		76890.9	114086.5
1	F142	F	NAD + Alcohol	2	70637.47	54988.69	86134.48		69929.7	94239.57

The complete raw data are given as an electronic file in Excel format and is attached online as part of the Supporting Information (see S2_File.xlsx).

References

1. Dona AC, Jiménez B, Schäfer H, Humpfer E, Spraul M, Lewis MR, et al. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal Chem*. 2014; 86(19):9887–9894.
2. Ellinger JJ, Chylla RA, Ulrich EL, Markley JL. Databases and Software for NMR-based metabolomics. *Curr Metabolomics*. 2013; 1:28–40.
3. Krojanski HG, Lambert J, Gerikalan Y, Suter D, Hergenröder R. Microslot NMR probe for metabolomics studies. *Anal Chem*. 2008; 80:8668-8672.
4. Smilde AKM, Van der Werf M, Bijlsma S, Van der Werff-van der Vat BJC, Jellema RH. Fusion of mass spectrometry-based metabolomics data. *Anal Chem*. 2005; 77:6729–6736.
5. ThermoFisher Scientific. Infinity™ Uric Acid Liquid Stable Reagent. 2012:1–2. Available: <https://www.thermofisher.com/order/catalog/product/TR24321>.

REFERENCES

- Álvarez-Sánchez, B., Priego-Capote, F. & De Castro, M.D. 2010. Metabolomics analysis I. selection of biological samples and practical aspects preceding sample preparation. *Trends in analytical chemistry*, 29(2):111–119.
- Antti, H., Bollard, M.E., Ebbels, T., Keun, H., Lindon, J.C., Nicholson, J.K. & Holmes, E. 2002. Batch statistical processing of ^1H NMR-derived urinary spectral data. *Journal of chemometrics*, 16(8–10):461–468.
- Beger, R.D., Dunn, W., Schmidt, M.A., Gross, S.S., Kirwan, J.A., Cascante, M., Brennan, L., Wishart, D.S., Oresic, M., Hankemeier, T., Broadhurst, D.I., Lane, A.N., Suhre, K., Kastenmüller, G., Sumner, S.J., Thiele, I., Fiehn, O. & Kaddurah-Daouk, R. 2016. Metabolomics enables precision medicine: “a white paper, community perspective”. *Metabolomics*, 12(10):149. doi:10.1007/s11306-016-1094-6.
- Bernini, P., Bertini, I., Luchinat, C., Nincheri, P., Staderini, S. & Turano, P. 2011. Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. *Journal of biomolecular NMR*, 49(3–4):231–243.
- Bouatra, S., Aziat, F., Mandal, R., Guo, A.C., Wilson, M.R., Knox, C., Bjorndahl, T.C., Krishnamurthy, R., Saleem, F., Liu, P., Dame, Z.T., Poelzer, J., Huynh, J., Yallou, F.S., Psychogios, N., Dong, E., Bogumil, R., Roehring, C. & Wishart, D.S. 2013. The human urine metabolome. *PLOS ONE*, 8(9):e73076. doi:10.1371/journal.pone.0073076.
- Chalmers, R.A., Healy, M.J., Lawson, A.M., Hart, J.T. & Watts, R.W. 1976. Urinary organic acids in man. III. quantitative ranges and patterns of excretion in a normal population. *Clinical chemistry*, 22(8):1292–1298.
- Coggon, D., Rose, G. & Barker, D.J.P. 1997. Longitudinal studies. (*In* Epidemiology for the uninitiated. 4th ed. London: BMJ Publishing Group. Chapter 7). <http://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/7-longitudinal-studies> Date of access: 14 Jan. 2017.
- Dunn, W.B., Broadhurst, D., Ellis, D.I., Brown, M., Halsall, A., O'Hagan, S., Spasic, I., Tseng, A. & Kell, D.B. 2008. A GC–TOF–MS study of the stability of serum and urine metabolomes during the UK biobank sample collection and preparation protocols. *International journal of epidemiology*, 37(suppl.1):i23–i30.

- Emwas, A.H., Luchinat, C., Turano, P., Tenori, L., Roy, R., Salek, R.M., Ryan, D., Merzaban, J.S., Kaddurah-Daouk, R., Zeri, A.C. & Gowda, G.N. 2015. Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review. *Metabolomics*, 11(4):872–894.
- Gibney, M.J., Walsh, M., Brennan, L., Roche, H.M., German, B. & Van Ommen, B. 2005. Metabolomics in human nutrition: opportunities and challenges. *American journal of clinical nutrition*, 82(3):497–503.
- Gika, H.G., Theodoridis, G.A., Wingate, J.E. & Wilson, I.D.J. 2007. Within-day reproducibility of an HPLC–MS-based method for metabonomic analysis: application to human urine. *Journal of proteome research*, 6(8):3291–3303.
- Gika, H.G., Macpherson, E., Theodoridis, G. & Wilson, I.D.J. 2008. Evaluation of the repeatability of ultra-performance liquid chromatography–TOF–MS for global metabolic profiling of human urine samples. *Journal of chromatography B*, 871(2):299–305.
- Hendriks, M.M.W.B., Van Eeuwijk, F.A., Jellema, R.H., Westerhuis, J.A., Reijmers, T.H., Hoefsloot, H.C.J. & Smilde, A.K. 2011. Data-processing strategies for metabolomics studies. *Trends in analytical chemistry*, 30(10):1685–1698.
- Jansen, J.J., Hoefsloot, H.C., Boelens, H.F., Van der Greef, J. & Smilde, A.K. 2004. Analysis of longitudinal metabolomics data. *Bioinformatics*, 20(15):2438–2446.
- Kamleh, M.A., Timothy M.D., Ebbels, T.M.D., Spagou, K., Masson, P. & Want, E.J. 2012. Optimizing the use of quality control samples for signal drift correction in large-scale urine metabolic profiling studies. *Analytical chemistry*, 84:2670–2677.
- Lesaffre, E. & Verbeke, G. 2005. Clinical trials and intervention studies. (In Everitt, B.S. & Howell, D.C., eds. *Encyclopedia of statistics in behavioral science*. New York, NY: Wiley. v. 1. p. 247–249).
- Lindon, J.C., Holmes, E. & Nicholson, J.K. 2001. Pattern recognition methods and applications in biomedical magnetic resonance. *Progress in nuclear magnetic resonance spectroscopy*, 39(1):1–40.

- Markley, J.L., Brüschweiler, R., Edison, A.S., Eghbalnia, H.R., Powers, R., Raftery, D. & Wishart, D.S. 2017. The future of NMR-based metabolomics. *Current opinion in biotechnology*, 43:34–40.
- Mason, S.W. 2010. The metabolomics of acute alcohol abuse. Potchefstroom: NWU. (Dissertation – M.Sc.).
- Numminen, H., Syrjälä, M., Benthin, G., Kaste, M. & Hillbom, M. 2000. The effect of acute ingestion of a large dose of alcohol on the hemostatic system and its circadian variation. *Stroke*, 31(6):1269–1273.
- Pasikanti, K.K., Ho, P.C. & Chan, E.C. 2008. Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. *Journal of chromatography B*, 871(2):202–211.
- Rubingh, C.M., Bijlsma, S., Jellema, R.H., Overkamp, K.M., Van der Werf, M.J. & Smilde, A.K. 2009. Analyzing longitudinal microbial metabolomics data. *Journal of proteome research*, 8(9):4319–4327.
- Saude, E.J. & Sykes, B.D. 2007. Urine stability for metabolomic studies: effects of preparation and storage. *Metabolomics*, 3(1):19–27.
- Saude, E.J., Adamko, D., Rowe, B.H., Marrie, T. & Sykes, B.D. 2007. Variation of metabolites in normal human urine. *Metabolomics*, 3(4):439–451.
- Smilde, A.K., Jansen, J.J., Hoefsloot, H.C.J., Lamers, R.-J.A.N., Van der Greef, J. & Timmerman, M.E. 2005. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics*, 21(13):3043–3048.
- Smolinska, A., Blanchet, L., Buydens, L.M. & Wijmenga, S.S. 2012. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. *Analytica chimica acta*, 750:82–97.
- t'Kindt, R., Morreel, K., Deforce, D., Boerjan, W. & Van Bocxlaer, J. 2009. Joint GC–MS and LC–MS platforms for comprehensive plant metabolomics: repeatability and sample pre-treatment. *Journal of chromatography B*, 877(29):3572–3580.

- Verouden, M.P., Westerhuis, J.A., Van der Werf, M.J. & Smilde, A.K. 2009. Exploring the analysis of structured metabolomics data. *Chemometrics and intelligent laboratory systems*, 98(1):88–96.
- Want, E.J., Wilson, I.D., Gika, H., Theodoridis, G., Plumb, R.S., Shockcor, J., Holmes, E. & Nicholson, J.K. 2010. Global metabolic profiling procedures for urine using UPLC–MS. *Nature protocols*, 5(6):1005–1018.
- Wishart, D.S. 2008a. Metabolomics: applications to food science and nutrition research. *Trends in food science & technology*, 19(9):482–493.
- Wishart, D.S. 2008b. Quantitative metabolomics using NMR. *Trends in analytical chemistry*, 27(3):228–237.
- Yee, J.L. & Niemeier, D. 1996. Advantages and disadvantages: longitudinal vs. repeated cross-section surveys. *Project battelle*, 94:16–22.
- Zelena, E., Dunn, W.B., Broadhurst, D., Francis-McIntyre, S., Carroll, K.M., Begley, P., O'Hagan, S., Knowles, J.D., Halsall, A., Wilson, I.D. & Kell, D.B. 2009. Development of a robust and repeatable UPLC–MS method for the long-term metabolomic study of human serum. *Analytical chemistry*, 81(4):1357–1364.
- Zhang, S., Gowda, G.N., Ye, T. & Raftery, D. 2010. Advances in NMR-based biofluid analysis and metabolite profiling. *Analyst*, 135(7):1490–1498.

PART III GC–MS METABOLOMICS OF ACUTE ALCOHOL CONSUMPTION

CHAPTER 5 GC–MS METABOLOMICS OF ACUTE ALCOHOL CONSUMPTION

5.1 Metabolism: The main root of GC-based metabolomics

In a way, the fundamentals of metabolism (and by association metabolomics) date back to ancient and primitive versions of body fluid screening tests. Anecdotal indications are that, as early as 2000–1500 B.C., traditional Chinese doctors used ants to evaluate patients' urine to determine if samples contained high levels of glucose, thereby indicating diabetes. Others also tasted urine for sweetness to check for the same thing (Van der Greef & Smilde, 2005). Hippocrates' examination of urine in the fourth century B.C., documented in *Aphorisma*, describes many observations that can today be linked to certain diseases (Armstrong, 2007). During the Middle Ages, "urine charts" were created and used to evaluate urine samples based on colour, taste and smell to detect various medical conditions (Armstrong, 2007; Nicholson & Lindon, 2008). Today we know that changes in the smell or colour of urine are associated with changes in the concentration of chemical components, and dysregulation of biochemical pathways, which are indicators of metabolic diseases, or of perturbations related to diet or to some other exogenous cause.

The term "metabolism", from the Greek μεταβολή – metabolē (change), was used by Justus von Liebig (1803–1873), in 1842, when he developed a theory of metabolism to describe the use of chemistry in plant and animal physiology. In contrast, the term "metabolomics" was coined only less than 20 years ago, at the end of the 20th century, and the field of metabolomics only truly started to develop at the beginning of the 21st century. This can be seen from the exponential growth in the number of publications on the topic (Figure 5–1). The main root of metabolomics, however, is the strong foundation of metabolism (and perturbed metabolism) that was explored and systemized in the 20th century. This knowledge of metabolism set the scene for the development of the field of metabolomics.

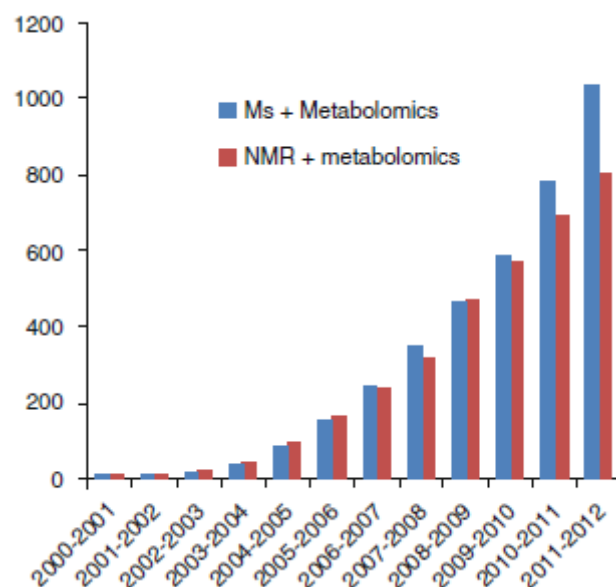


Figure 5-1: Number of NMR metabolomics and MS metabolomics publications [reproduced with permission from Emwas, 2015]. The literature review was conducted using Web of Knowledge (<http://apps.webofknowledge.com>) with the keywords “metabolomics and NMR” and “metabolomics and MS”.

A major focus in the field of metabolism in the 20th century was the identification and diagnosis of diseases, especially inborn errors of metabolism (IEMs) — an application that is still of key importance in present-day metabolomics. The research on IEMs was important for the development of metabolomics from two points of view: (1) the experiments on IEMs were based on good separation and identification techniques, of which GC and MS, as well as the combination of the two (GC–MS), were some of the foremost techniques used; and (2) IEMs are a class of genetic diseases characterized by disorders of the metabolism. These diseases are mostly caused by defects of single genes that code for enzymes necessary and responsible for certain metabolic pathways. Disruption of these pathways results in:

1. The production of secondary metabolites which are not always detectable under normal homeostasis.
2. The accumulation of toxic substances which could interfere with normal metabolic functions.
3. The reduced synthesis of essential metabolic compounds (Emwas *et al.*, 2013).

The first reference to the concept of IEMs was by the British physician Sir Archibald Garrod early in the 20th century, when he suggested that changes in metabolic pathways were caused by a single gene defect (historical review by Bearn & Miller, 1979; Van der Greef *et al.*, 2013). He is credited with identifying and describing some of the first IEMs, namely albinism, alkaptonuria, cystinuria, and pentosuria.

Around the same time, Otto Knut Olof Folin developed new analytical approaches to urine analysis in order to compare the composition of urine from normal and insane people (historical review by Simoni *et al.*, 2002). In 1905, he published 3 papers in the *American Journal of Physiology* in which he reported methods for the urinary analysis of the major non-protein nitrogen-containing compounds in urine, namely urea, ammonia, creatine, creatinine and uric acid. Unlike previous methods, these techniques were relatively specific and sensitive and only required small sample volumes. This work prompted many other researchers to study urine to gain understanding of normal and pathophysiological processes. According to E.V. McCollum, this development “marked a turning point in the history of biochemistry” (Simoni *et al.*, 2002).

As early as the late 1940s, Roger Williams and his associates used paper chromatography (which later developed into the widely applied thin-layer metabolite and carbohydrate screening tests) to show that characteristic metabolic patterns are reflected in an individual’s urine and saliva, and that certain patterns were associated with particular disease states (documented in a review by Gates & Sweeley, 1978; review by Novotny *et al.*, 2008). Although not called such at the time, this work by Williams and his associates is the first reference to the idea that an individual might have a “metabolic profile” that would be reflected in the constituents of their biological fluids. The concept, however, essentially lay dormant until the late 1960s when advances in gas and liquid chromatography introduced methods that were considerably less laborious, and of higher sensitivity and specificity, than these pioneering approaches.

Until the 1960s, most metabolic experiments were mainly qualitative. It was only following technological advancements during this time that highly sensitive and quantitative measurement of metabolic profiles, and the ability to study individual metabolites, became possible (Preti, 2005):

1. The discovery and extensive application of chromatographic separation methods, and the development of new sensitive ionization detectors, during the 1950s and 1960s, established chromatography as a quantitative analytical method for multi-component determinations (Novotny *et al.*, 2008).
2. The continuous development of MS methods and machinery during the 1960s and 1970s (after the initial development of MS more than a century ago in 1905 by J.J. Thomas – Scripps Metabolomics Science timeline), led MS to become a highly specific analytical tool for providing information on such details as accurate mass, isotope distribution patterns, structural elucidation and metabolite identification, and comparative concentrations of different metabolites in complex mixtures (Emwas, 2015).

However, on its own, GC provided neither sufficient specificity of detector response nor adequate chromatographic resolution to permit unequivocal identification of most peaks in

complex mixtures, complicating quantitative analyses (Gates & Sweeley, 1978). The solution came when the gas chromatograph could be coupled to a mass spectrometer, first demonstrated in 1959 by Gohlke, and subsequently by others (e.g. in 1961 by both Henneberg and Ebert) (Gates & Sweeley, 1978; Ryhage, 1964), which extended the usefulness of both instruments.

In 1966, Dalgliesh *et al.* used GC to separate and identify — using a series of straight-chain hydrocarbons as reference standards in order to express the retention times/peak positions of the peaks in the chromatograms of the samples as “methylene units” — a wide range of substances present in urine and tissue extracts chromatographically. They then confirmed the identities of all the substances, identified in this way, by coupling GC to MS in order to also characterize the substances structurally. By doing so, the Dalgliesh group were among the first researchers to show the power of GC–MS in the evaluation of metabolic patterns. They demonstrated that it is possible to separate, identify and measure, on the same chromatogram, derivatives of a variety of trace organic compounds present in urine and tissue extracts, such as citric acid cycle intermediates, steroids and metabolites of biological amines. Their findings were published as a keynote paper in *Biochemical Journal*, entitled “A gas-liquid-chromatographic procedure for separating a wide range of metabolites occurring in urine or tissue extracts” (Dalgliesh *et al.*, 1966).

The technological developments and the subsequent biological discoveries also gave rise to new terminology. The terms “metabolite pattern”, used by the group of S.K. Wadman in The Netherlands (Duran *et al.*, 1978), and “metabolic profile”, used by the Horning group in the United States (Horning & Horning, 1971), became established through the development of GC–MS procedures, and are used to define or describe the metabolic patterns of groups of metabolically or analytically related metabolites, such as urinary steroids, urinary organic and amino acids, and neutral and acidic urinary drug metabolites (Horning & Horning, 1971). These pioneering groups — through their recognition that these metabolite patterns could be useful for characterising both normal and pathological states, for studies of drug metabolism and the effects of drugs on human metabolism, and for human developmental studies — led the way for the development and use of GC–MS methods to monitor urine metabolites during the 1970s (Griffiths & Wang, 2009).

A next major contribution came in 1980 when Tanaka *et al.* used a GC method similar to that used by the Dalgliesh group to characterize 155 metabolic standards (metabolic intermediates present in normal urine as well as unusual metabolites present in the urine of people suffering from organic acidurias) in terms of their “methylene units”. The lists compiled by this group made it possible to identify and diagnose a range of organic acidurias using GC alone. Their results, lists and practical implementation of the method to diagnose diseases were published in

two widely used papers in *Clinical Chemistry* (Tanaka, Hine, *et al.*, 1980; Tanaka, West-Dull, *et al.*, 1980). This method was especially useful at the time, because, although available, GC–MS instrumentation was expensive and required much more specialized training and technical expertise than GC alone, and therefore not many medical screening centres were equipped with these instruments. However, in time, using GC techniques for separation purposes together with MS methods for identification and determining structures, GC–MS became one of the most powerful separation methods for the analysis of complex biological samples, providing qualitative and quantitative information on volatile compounds (or compounds that can be made volatile through chemical derivatization).

It is evident that the metabolic studies of this time focused mostly on identifying and measuring individual metabolites (which can be seen as “targeted studies”) to diagnose and identify biomarkers of disease, especially IEMs. Serious to fatal consequences are associated with many IEMs if not diagnosed and treated early, and therefore a rapid, sensitive and general method of detection is ideal (Lindon *et al.*, 2000). GC–MS provides highly reproducible chromatographic separation and fragmentation, making it highly suitable for metabolite identification. Therefore, despite the various limitations inherent to GC–MS (Emwas, 2015) (which will be discussed in the following section), it remains the preferred method of analysis for targeted studies where the primary aims are disease diagnosis and tracking of disease progression or treatment outcome.

The advances in GC–MS technology and instrumentation in the 1960s and 1970s greatly enhanced the qualitative and quantitative detection and diagnosis of IEMs, and led to a dramatic rise in the number of these diseases being identified by an increasing number of international clinical chemists and groups (Duran *et al.*, 1994; Eldjarn *et al.*, 1975; Jellum, 1977; Saudubray *et al.*, 2006; Wadman *et al.*, 1971). At least 298 commonly detectable disorders and their diagnostic methods are described in a scholarly textbook on IEMs, and are grouped into 35 chapters according to the type of condition (Blau *et al.*, 2008). These developments demanded a new level of quality control — diagnostic methods, which included GC–MS as an important technology, had to be organized, and technical and professional staff had to be suitably trained in order to serve potential patients at the national and international level.

The ERNDIM (European Research Network for evaluation and improvement of screening, Diagnosis and treatment of Inherited disorders of Metabolism) foundation was established in 1994 to meet these challenges (ERNDIM, 2017; Fowler *et al.*, 2008; Peters *et al.*, 2008). ERNDIM organizes European Quality Assurance (EQA) schemes for qualitative and quantitative assays of organic acids, amino acids and other special metabolites according to accepted norms on a mainly Europe-wide scale. Several laboratories from outside Europe also participate, however, including the Potchefstroom Laboratory for Inborn Errors of Metabolism

(PLIEM) (www.pliem.co.za). The organic acids in the urine samples used in the present study were extracted and analysed according to the standard operational procedure (SOP) of the PLIEM laboratory, and will be elaborated on below.

Thus, by the end of the 20th century, a vast field of knowledge on metabolism and its associated inherited disorders existed, and was still expanding. However, research on, and the mapping of, the human genome introduced a next, and fundamental, new era — the advent of genomics and the subsequent new fields of omics technologies, which again gave rise to new terminologies. The term “metabolome” was first used by S.G. Oliver in 1998 in a publication titled “Systematic functional analysis of the yeast genome”, published in *Trends in Biotechnology* (Oliver *et al.*, 1998), and the term “metabolomics” by O. Fiehn in his paper “Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks”, published in *Comparative and Functional Genomics* in 2001 (Fiehn, 2001). The term “metabolome” is defined as “the entire complement of all the small-molecular weight molecules (metabolites in cells, body fluids, tissues, etc.)” (Van der Greef & Smilde, 2005), and “metabolomics” as “the quantitative measurement of the entire complement of low molecular weight metabolites in an organism’s cells at a specific time under specific environmental conditions” (Holmes & Antti, 2002). The word “metabonomics” was introduced in 1999 by Nicholson *et al.* in their publication “Metabonomics: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data”, and defined as “the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification” (Nicholson *et al.*, 1999).

From this overview, it can therefore be said that the metabolic studies of the 20th century (especially those using GC–MS) provided a solid root and firm directive on which present-day metabolomics could be built.

5.2 Sources of variation and quality control in GC-metabolomics

Some of the main aspects of metabolomics research are:

- The precision of the quality of the data used to assess and optimize the metabolic phenomenon under investigation.
- To make use of minimum reporting standards for data analysis. This involves an unambiguous description of the study and its execution, including the subjects, sample collection and preparation, experimental analysis, quality control, metabolite identification, and data pre-processing. These standards do not aim to prescribe how experiments are to be carried out, but rather to ensure that the data and new knowledge gained from

metabolomics investigations can be shared and re-used by other researchers in useful ways (Fiehn *et al.*, 2007; Goodacre *et al.*, 2007; Sumner *et al.*, 2007).

Quality control procedures are fundamental components required for good laboratory practices, and will now be addressed.

It is clear from the foregoing that GC–MS is an excellent method for the quantification and identification of metabolites. Since the pioneering views on the potential of metabolomics research (Fiehn, 2002), GC–MS has become one of the most frequently used methods for metabolomics investigations due to its high sensitivity, peak resolution, reproducibility, low cost and ease of instrumental use (Dunn & Ellis, 2005; Jiang *et al.*, 2010; Kanani *et al.*, 2008; Kopka, 2006; Pasikanti *et al.*, 2008; Quéro *et al.*, 2014; Want *et al.*, 2007; Xu *et al.*, 2010). Indeed, GC–MS has been referred to as the “gold standard” for metabolomics due to the high separation power of GC and better deconvolution algorithms that are available (compared to other platforms) (Dunn & Ellis, 2005; Harrigan & Goodacre, 2003; Weckwerth & Fiehn, 2002). Using GC–MS in metabolomics, however, is not without its challenges and limitations — such as that only volatile compounds or compounds that can be made volatile by chemical derivatization can be analysed.

A typical GC–MS metabolomics experiment involves several successive experimental procedures, and practically every analytical step in such a GC–MS experiment could lead to variation in the data (Fukusaki & Kobayashi, 2005). It is essential that the sources of variation be well known and defined (Kanani *et al.*, 2008). Only then can the fluctuations in metabolite levels and the results obtained be considered to be reliable and useful, and can the changes detected in metabolic profiles be ascribed to the important biological variation — and a response to the intervention or disease being studied — as opposed to non-biological analytical (or other) variation (t’Kindt *et al.*, 2009).

Within a typical metabolomics study there are two types of variation: technical and biological. Technical (or analytical or experiment-originating) variation refers to potential biases that can arise from the analytical method (such as sampling, sample work-up, analytical errors and instrumental noise) (Hendriks *et al.*, 2011). Biological variation can be categorized into three types (Van den Berg *et al.*, 2006; Want *et al.*, 2010):

1. Induced/relevant – variation caused by a pathological condition of interest or by experimental conditions designed to induce a desired biological effect (such as the ingestion of alcohol as examined in this study).
2. Non-relevant – intra-individual and inter-individual variation that occurs under normal physiological conditions (and not due to the pathological or experimental conditions), that can be monitored, accounted for and, to an extent, controlled.

3. Uninduced – variations in certain metabolite concentrations under identical experimental conditions.

Thus, the total variation within the original data generated in a metabolomics experiment consists of the sum of all induced/relevant, non-relevant and uninduced biological variations, as well as all technical variations. Data acquisition — including sample collection, sample pre-treatment and analysis — should be aimed at minimising the sources and effects of technical variation, so that the overall variability in a well-designed metabolomics study is primarily related to biological origin (Hendriks *et al.*, 2011; Shurubor *et al.*, 2007; Van den Berg *et al.*, 2006). In order to understand, assess the effect of and account for (or minimize) sources of technical variation, many GC–MS study designs include method validation and quality assurance procedures. This includes using validated experimental methods and well-defined SOPs, as well as the addition of internal standards, standardized mixtures and/or QC samples to the experimental analysis protocol (Gika *et al.*, 2007; Hendriks *et al.*, 2011; Pasikanti *et al.*, 2008; Van Batenburg *et al.*, 2011; Xu *et al.*, 2010; Zelena *et al.*, 2009). Method validation helps to ensure reliable data production and to demonstrate that a particular method, used for quantitative measurement of analytes in a given biological matrix, is reliable and reproducible for the intended use (Sumner *et al.*, 2007).

Tables 5–1 to 5–7 list (with references) the analytical procedures for the GC–MS analysis of organic acids in urine samples, the sources of variation that can possibly be introduced at each step, and the quality control measures that can be taken to minimize or eliminate the listed sources of variation, with the focus placed on aspects that are relevant to the present investigation.

Tables 5-1 to 5-7: Analytical procedures for the GC–MS analysis of organic acids in urine samples, the sources of variation associated with each step, and the quality control measures that can be taken to minimize or eliminate these sources of variation.

Table 5-1: Sampling procedures

	Sources of variation	QC procedures	References
General	Demographic and environmental factors like gender, age and ethnicity.	Limit non-relevant biological variation, and other factors that could affect the baseline metabolite profile as far as possible by using only samples from a well-defined group that is as homogeneous as possible, and collected under controlled environmental conditions (diet, lifestyle, etc.) in order to obtain valuable reliable information. This, however, often requires an increased number of samples in order to sufficiently represent different groups, which is not always practically feasible. Alternatively, the use of proper controls can help to minimize this source of variation.	Dunn & Ellis, 2005 Pasikanti <i>et al.</i> , 2008 Zelena <i>et al.</i> , 2009
Collection	Contamination and degradation.	First-void urine samples are preferred over spot urine samples, since the influence of lifestyle factors (such as diet) should be relatively minimized and normalized in these samples. Should be rapid, follow strict SOPs and require the minimum number of steps. Careless sampling may cause significant experimental variation, which could surpass biological variation. Sample details (i.e. volume, time of day, metabolic state, diet, medication, etc.), as well as details on the subjects, should be noted, since it allows for a better biological interpretation of the data generated. Although their use is not recommended, since they introduce additional peaks and may cause changes in the pH and ionic strength of samples, the addition of any preservatives or other chemicals should be noted.	Bernini <i>et al.</i> , 2011 Castle <i>et al.</i> , 2006 Fiehn <i>et al.</i> , 2007 Fukusaki & Kobayashi, 2005 Pasikanti <i>et al.</i> , 2008 Saude <i>et al.</i> , 2007 Xu <i>et al.</i> , 2010
Handling and transport	Bacterial contamination and sample degradation. Changes in the concentrations of metabolites due to residual enzymatic activity in samples.	Conditions should be minimal and well documented, since sample stability is essential for accurate, valid and reproducible results. Appropriate sample handling methods ensure that the analysed sample is representative as much as possible of the metabolome before collection (i.e. as close as possible to the freshly collected sample). During transport, samples should preferably be kept on ice or at 4°C to avoid degradation. Mild pre-centrifugation combined with filtration after sample collection should quench bacterial/enzymatic activity and minimize contamination.	Bernini <i>et al.</i> , 2011 Dunn <i>et al.</i> , 2008 Xu <i>et al.</i> , 2010
Storage	Compound instability and degradation.	Samples should either be analysed immediately, or otherwise frozen and stored at the lowest possible temperature, preferably –80°C. Samples should undergo as few as possible freeze–thaw cycles (defrosting a stored sample for analysis and then refreezing the same sample for future analyses) in order to maintain their integrity. Samples required for multiple analyses should be divided into various aliquots to avoid multiple freeze–thaw cycles.	Dunn & Ellis, 2005 Gika <i>et al.</i> , 2007 Pasikanti <i>et al.</i> , 2008 Teahan <i>et al.</i> , 2006 Xu <i>et al.</i> , 2010

Table 5-2: Preparation of samples for GC–MS analysis

	Sources of variation	QC procedures	References
Extraction	Extensive, time-consuming and repetitive sample preparation, potentially leading to experimental error. Inaccurate measurements of reagents and vortex times, insufficient time allowed for extraction, varying amount of supernatant recovery after centrifugation, varying extent of sample drying, and inconsistent experimental temperatures.	These variables should be evaluated prior to analysis and strictly controlled during analysis. Follow validated methods and SOPs as far as possible to obtain an optimal and uniform degree of extraction, and minimal variation. Variations can be monitored and accounted for through the use of internal standards, since they should affect all metabolites in a sample in the same way.	Emwas, 2015 Kanani <i>et al.</i> , 2008 Xu <i>et al.</i> , 2010
Derivatization	Formation of multiple silyl derivatives and multiple peaks for a single compound, due to inconsistent or incomplete derivatization of compounds, as well as the formation of artifacts or unexpected derivatives and metabolite conversions, affect overall reproducibility and complicates metabolite identification and quantification. Incomplete derivatization due to incorrect derivatization conditions (i.e. insufficient derivatization reagent, time and/or temperature) and/or steric hindrance in compounds with multiple functional groups. Reversibility of silylation reaction in the presence of water. Degradation of thermally labile metabolites during incubation period where samples are heated for a defined period of time. Stability of trimethylsilyl (TMS) derivatives. Reagents cause damage to GC column. High concentrations of TMS by-products can saturate the MS detector.	The analyst should be aware of the possibility of artifacts and multi-peak formation, and care should be exercised when interpreting the results. Allowing sufficient reaction time at an appropriate temperature, and ensuring that the amount of derivatization reagents added is in excess compared to the sample compounds that need to be derivatized should allow for complete derivatization of sample compounds, minimising variation and multiple peak formation. However, longer derivatization times could potentially lead to degradation of unstable compounds. Therefore, the derivatization process should be carefully monitored and optimized (e.g. by following SOPs) with regard to reaction mechanisms, reaction times and temperatures, as well as choice and quantity of solvents, derivatization reagents and catalysts. Derivatized samples should not be stored at room temperature for long periods of time to avoid sample degradation and changes in sample composition. Derivatized samples must preferably be kept in tightly closed vials at 4°C or lower on the autosampler between derivatization and analysis. Sample stability (which is highly dependent on the sample environment and conditions), as well as damage to the GC column, can be monitored with the addition of QC samples to the analytical runs. The MS detector could be turned off until all TMS by-products have eluted, but should be carefully controlled since it could result in fast eluting compounds not being detected.	Bedair & Summer, 2008 Dunn & Ellis, 2005 Dunn <i>et al.</i> , 2005 Halket <i>et al.</i> , 2005 Kanani <i>et al.</i> , 2008 Kouremenos <i>et al.</i> , 2010 Little, 1999 Pasikanti <i>et al.</i> , 2008 Quéro <i>et al.</i> , 2014 Villas-Bôas <i>et al.</i> , 2005 Xu <i>et al.</i> , 2010

Table 5-3: GC separation of sample compounds

	Sources of variation	QC procedures	References
Analytical platform	Instrumental noise. Overall performance.	Standardized mixtures (solutions prepared from pure commercial standards containing a mixture of compounds of known concentrations) should be injected at the beginning, middle and end of an analytical run. The results are used to assess chromatographic and detector performance across the GC–MS run, RT stability, peak shape, signal intensity, mass accuracy, and detector response, to detect any dramatic changes to the system, and to detect poor quality data sets that can then be removed from subsequent analysis. QC samples (injected at regular intervals throughout the analytical run) are also used to monitor the stability and performance of the system with regard to RT, detector response and mass accuracy. QC samples are expected to cluster closely together and show little to no variation or time-related trends when applying statistical techniques such as PCA to the data. QCs showing high variability should be easily detectable, leading to questions regarding the reliability of the rest of the data from that run.	Gika <i>et al.</i> , 2007 Gika <i>et al.</i> , 2008 Pasikanti <i>et al.</i> , 2008 t'Kindt <i>et al.</i> , 2009 Van Batenburg <i>et al.</i> , 2011 Want <i>et al.</i> , 2010 Zelena <i>et al.</i> , 2009
Sample injection	Inconsistent volume. Split or splitless injection. Erratic and non-reproducible results are obtained from using stainless steel injector ports for derivatized compounds. Sample stability.	Small volumes of derivatized sample (approximately 1 µL) are injected into the GC column by split or splitless techniques. Splitless injection introduces a larger amount of sample to the column, but a split system is preferred when the detector is sensitive to trace amounts of analytes and if the amount of sample can overload the column. Glass injector ports should be used for injecting derivatized compounds to obtain more reproducible results. Multiple sample injections from the same vial decreases sample stability and increases analytical variability. Single sample injections per vial reduce variability.	Douglas, 2004 Pasikanti <i>et al.</i> , 2008 Quéro <i>et al.</i> , 2014 Van Batenburg <i>et al.</i> , 2011
Injection port septum	Leaking or worn septum could introduce contaminants into sample, which could alter its characteristics.	Septum should be replaced every 100–200 injections, and more regularly if higher injection port temperatures are used.	Douglas, 2004
Column	Contamination and deterioration (due to sample components that remain in the injector or column) cause drift. Modification or replacement can cause retention time (RT) shifts. Variation in first samples analysed.	QC samples can be used to monitor and correct for drift. GC columns should be replaced with identical columns to maintain separation repeatability. Conditioning of the column with the injection of multiple QC samples at the start of the run equilibrates the system to the samples, increases the stability of the system, and the repeatability of the analysis, by minimising RT disturbance and drift in the subsequent analyses.	Gika <i>et al.</i> , 2007 Gika <i>et al.</i> , 2008 Halket <i>et al.</i> , 2005 Kamleh <i>et al.</i> , 2012 Van Batenburg <i>et al.</i> , 2011 Zelena <i>et al.</i> , 2009

Temperature	Sample instability may occur at high temperatures, which might lead to the formation of multiple peaks for single compounds. Too low an injection port temperature may result in incomplete volatilization of sample and broad, poorly separated peaks or no peaks.	Derivatization increases the thermal stability and volatility of metabolites, which prevents their degradation at the high temperature required for GC analysis. The injection port temperature should be optimized and maintained according to SOPs to ensure complete sample volatilization.	Douglas, 2004 Dunn & Ellis, 2005 Halket <i>et al.</i>, 2005 Pasikanti <i>et al.</i>, 2008 Van Batenburg <i>et al.</i>, 2011 Xu <i>et al.</i>, 2010
Separation	Late-eluting compounds from previous runs. Some compounds become attached to the column. Multiple peaks for single compounds due to formation of derivatization artifacts leading to overlapping peaks. Column properties. Carrier gas type and velocity. Oven temperature programme.	Blank samples may be run between experimental samples to prevent compounds from previous runs eluting with those from subsequent runs. Derivatization decreases the polarity of metabolites, which prevents interactions between the samples and the GC column, and limits the number of metabolites that become attached to the column. It also improves chromatographic behaviour (peak symmetry and RT stability) and separation of compounds. Long analysis time (up to 60 minutes per sample) is often required for satisfactory separation of overlapping peaks. Optimising various GC parameters maximizes metabolite separation before MS detection: (1) Column length and stationary phase polarity influence separation efficiency. Capillary GC columns made of fused silica (such as DB-5MS) are most commonly used and provide higher chromatographic resolution. (2) Column flow rate between 0.8 and 2 mL/min is commonly used. Higher carrier gas flow rates results in faster analysis times, but poorer separation between the analytes. (3) Typical oven temperatures range from 40 to 325°C. Higher temperatures result in faster analysis times, but often at the cost of sufficient separation.	Dunn & Ellis, 2005 Gika <i>et al.</i>, 2007 Halket <i>et al.</i>, 2005 Pasikanti <i>et al.</i>, 2008 Van Batenburg <i>et al.</i>, 2011 Xu <i>et al.</i>, 2010

Table 5-4: MS analysis of separated compounds

	Sources of variation	QC procedures	References
Fragmentation	Extensive fragmentation caused by electron-impact ionization complicates metabolite identification. Fragmentation patterns can be affected by changes in MS conditions, such as variation in ionization assembly, the electron multiplier or alignment of the mass spectrometer through various runs.	The energy supplied during EI is kept constant, enabling reproducible mass spectra to be obtained, which facilitates library searches and metabolite identification. MS operating conditions should be regularly monitored, and any results obtained under different conditions should be interpreted with care. Ideally, only the results from samples analysed under the same conditions should be used and compared.	Halket <i>et al.</i> , 2005 Kanani <i>et al.</i> , 2008 Want <i>et al.</i> , 2007
Interior MS pressure	Too high an interior pressure could cause fragments to collide and form new particles, which complicates compound identification.	The interior MS pressure should be optimized and kept low to prevent fragment collisions.	Douglas, 2004
Ion source	Contamination reduces ionization efficiency (due to co-elution), and causes altered signal intensity and drift.	QC samples are used to monitor and correct variation and drift due to contamination. Longer separation times prevent co-elution of contaminants with target metabolites.	Fukusaki & Kobayashi, 2005 Halket <i>et al.</i> , 2005 Kamleh <i>et al.</i> , 2012
Detector response	Poor or inconsistent detection of low-intensity peaks leading to larger signal variability, as well as zero and outlier values. Too low a scan speed can lead to peak skewing, which can influence quantification. No MS detector exists that can detect all metabolites and no single ionization methods is able to ionize all metabolites.	The analyst should be aware of the poorer detector response for low-intensity peaks, and care must be taken when interpreting such results. Software such as AMDIS can be used to correct for peak skewing, but care must be taken when performing such corrections. QC samples are used to monitor and correct variation and drift due to changes in the detector response. The number and class of detected metabolites are highly dependent on the choice of ionization method. Different methods should be used independently to maximize the number of metabolites detected.	Emwas, 2015 Gika <i>et al.</i> , 2007 Gika <i>et al.</i> , 2008 Halket <i>et al.</i> , 2005 Kamleh <i>et al.</i> , 2012 Van Batenburg <i>et al.</i> , 2011
General	Changes in environmental temperature and separation conditions can cause variation in recorded metabolite response.	QC samples are used to monitor and correct general variation in metabolite responses caused by such changes.	Kamleh <i>et al.</i> , 2012
Analysis time	Sample instability, product formation and degradation resulting in time-related trends. Stability of analysis for the duration of the run.	Samples should be analysed as soon as possible after preparation, and prepared samples should be kept at 4°C prior to analysis to avoid degradation of unstable sample components. QC samples are used to investigate the analytical variability over the time course of the analysis and are therefore an indication of the reliability of the data.	Gika <i>et al.</i> , 2007 Gika <i>et al.</i> , 2008 Pasikanti <i>et al.</i> , 2008

Table 5-5: Analysis of GC–MS spectra

	Sources of variation	QC procedures	References
Deconvolution	Peak-picking errors leading to signal variability.	Results should be carefully monitored and verified in order to avoid false positive and false negative results. Deconvolution software (such as AMDIS) can detect co-eluting peaks with peak apexes separated by less than 1 s, thereby improving the resolution of GC–MS and eliminating the need for complete chromatographic separation.	Dunn & Ellis, 2005 Quéro <i>et al.</i> , 2014
Quantification	External calibration or response ratio. Peak overlapping due to co-eluting metabolites, which complicates peak detection and quantification.	Since external calibration is labour intensive and requires standard solutions of all metabolites to be quantified, quantification is more frequently performed by calculating the response ratio, which expresses a metabolite's concentration relative to that of the known amount of an added internal standard. The pure chromatograms derived from deconvolution of overlapped data can be used to calculate relative metabolite concentrations.	Dunn & Ellis, 2005 Halket <i>et al.</i> , 2005 Hendriks <i>et al.</i> , 2011
Identification	Extensive fragmentation resulting from electron-impact ionization (EI) causes disappearance of molecular ions which hinders metabolite identification. Incomplete GC separation, chromatographic drift (due to variations in GC operational conditions), metabolites with very similar chemical structures, the large number of unknown peaks, peaks that result from column bleeding, and multiple derivative peaks all complicate metabolite identification. GC–MS is not general (only able to analyse small volatile compound or compounds that can be made volatile through derivatization), which limits its use for global metabolic profiling.	Chemical ionization can be used as an alternative to EI, since it is much less energetic and therefore produces more-stable ions and less fragmentation, simplifying the identification of the molecular ion, which aids metabolite identification. RT indices and peak alignment software are used to align metabolic profiles when chromatographic drift occurs. Derivatization enhances sensitivity and selectivity, and facilitates structural elucidation. Metabolite databases (such as AMDIS and NIST) aid in characterisation and identification of unknown metabolites. RT or retention index and mass spectra of sample peaks are compared with those of pure compounds analysed under the same conditions. However, since metabolite identification is so vital for the conversion of raw data into biological context, results preferably need to be verified or validated to ensure their integrity. Different MS methods need to be employed for the detection of different classes of metabolites in order to obtain a comprehensive global metabolic profile. The pure MS spectra derived from deconvolution of overlapped data can be used for metabolite identification.	Dunn & Ellis, 2005 Dunn <i>et al.</i> , 2008 Emwas, 2015 Halket <i>et al.</i> , 2005 Hendriks <i>et al.</i> , 2011 Kanani <i>et al.</i> , 2008 Pasikanti <i>et al.</i> , 2008 Quéro <i>et al.</i> , 2014 Sumner <i>et al.</i> , 2007 Want <i>et al.</i> , 2007 Xu <i>et al.</i> , 2010

Table 5-6: Analysing samples in multiple batches — batch effect

	Sources of variation	QC procedures	References
Batch length	Too large batches lead to drift in instrumental performance over time and affect data quality (sensitivity and RT). Too small batches negatively impact statistical validity.	Samples should be split into batches of optimized lengths — small enough to maintain technical repeatability of results, and large enough for effective statistical analysis, an aspect which is highly dependent on sample numbers. QC samples can be used to detect the presence of time-related machine/signal drift.	Kamleh <i>et al.</i> , 2012 Van Batenburg <i>et al.</i> , 2011 Zelena <i>et al.</i> , 2009
Intra- and inter-batch variation	Instrumental maintenance, and modification of replacement of column leading to signal drift.	Appropriate instrumental maintenance (to minimize contamination of the column or ion source) should be performed between separate batches to avoid disruption of analytical runs and the possibility of intra-batch variation. QC samples, included at regular intervals in the analytical runs, are used to monitor intra-batch repeatability and to correct for any observed batch effect. By using the same QC sample in all sample batches and comparing their results across all batches, inter-batch repeatability can also be assessed.	Hendriks <i>et al.</i> , 2011 Kamleh <i>et al.</i> , 2012 Zelena <i>et al.</i> , 2009
Re-equilibration time between batches	Too long times unnecessarily lengthen total analysis time. Too short times do not allow the chromatographic phase to re-equilibrate with the starting mobile phase, which could affect early-eluting peaks from next batch.	Experiment with different re-equilibration times to find the optimal compromise. Use reliable and tested GM–MS analysis programmes and SOPs for optimal results.	Gika <i>et al.</i> , 2007

Table 5-7: Data analysis and processing

	Sources of variation	QC procedures	References
General	Peak-picking errors or misalignment (due to baseline drift, RT shifts or selection of inappropriate RT windows, instrumental noise, and artifact formation) leading to signal variability, zero and outlier values. Over-optimistic or false data interpretation due to incorrect determination of the statistical significance of a model.	Input parameters of signal-alignment software should be carefully optimized and results should be double-checked for errors to avoid false positive and false negative results. Permutation tests reveal whether the classifications or predictions obtained from statistical analyses differ from more chance results, and are therefore used to determine the statistical significance of models.	Gika <i>et al.</i> , 2007 Gika <i>et al.</i> , 2008 Hendriks <i>et al.</i> , 2011 Kamleh <i>et al.</i> , 2012 Pasikanti <i>et al.</i> , 2008

5.3 From GC–MS data to biology

Being able to analyse biological samples precisely and accurately and obtain analytically reliable and repeatable data, inter- and intra-batch precision, and statistically significant results, is an important part of any metabolomics study. From only a cursory inspection of the content of Tables 5–1 to 5–7, it is evident that the sources of variation associated with GC–MS metabolomics experiments are many. This has resulted in the need for the development of extensive procedures for analytical precision (Jonsson *et al.*, 2005), different procedures for data analysis to improve biological information (Van den Berg *et al.*, 2006), initiatives to systematize metabolomics standards (Fiehn *et al.*, 2007), guidelines for reporting metabolomics data (Sumner *et al.*, 2007), as well as to define measurement errors and figures of merit (Van Batenburg *et al.*, 2011). However, despite these developments, the literature reporting metabolomics studies is fraught with heterogeneity with regard to the levels of reporting their details. Also, unspecified concepts and terminologies, emanating from traditional jargon specific to certain biological disciplines, are frequently encountered. Within this wilderness, and in the increasing digitisation of the world, “biology has become a data intensive field of disparate methods, with images, sequence reads and spectra, to name only a few, all being acquired by the droves” (Rocca-Serra *et al.*, 2016). However, if this magnitude of data and results obtained from metabolomics studies cannot be biologically interpreted correctly and linked to metabolism in a valuable way, they are rather useless. Additionally, the vast field of metabolism and metabolic pathways, together with the extensive biological diversity that exists, further complicates the biological interpretation of metabolomics data. Founded biological insights and knowledge are necessary for more than a presumed biological interpretation of metabolomics data — a lesson that Nobel laureate and leading chemist of the 20th century, Linus Pauling, had to learn the hard way:

In 1971 a most remarkable and innovative contribution of the time was made by Linus Pauling and his co-authors, and was published in *The Proceedings of the National Academy of Sciences*. In their paper, titled “Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography” (Pauling *et al.*, 1971), they proposed that, by placing a person on a defined diet consisting almost entirely of small molecules, intestinal flora (and therefore any influence they might have on the urine composition) are eliminated due to starvation. By doing so, the physical condition of a person can be standardized and the composition of his/her body fluids can be made constant after a few days. Through the use of this method and gas–liquid partition chromatography, they qualitatively identified many substances in breath and in urine vapour samples, which would “permit the quantitative determination of about 250 substances in a sample of breath, and of about 280 substances in a sample of urine vapor.” They also mention the difference between these

samples and samples taken from people not on the confined diet, implying the presence of biological variability when intestinal flora is present.

This highly notable paper describes the diagnostic and intervention potential of metabolite information, and, by stating “quantitation is achieved by a computer program that integrates the signals for the successive peaks, the integration being performed from minimum to minimum”, it directed towards metabolomics technology that only emerged three decades later, and unintentionally alerted to the multidisciplinary prerequisite nature of metabolomics research, which should involve biologists, chemists and statisticians. However, by not being a biologist, Pauling had a distorted view on the physiological importance of the gut microbiome. This was pointed out by Winitz *et al.*, who reported that a glucose-based diet decreases but does not entirely eliminate intestinal flora. Accordingly, it cannot be used to standardize a person or make the composition of their body fluids constant, and as a result this method does not eliminate biological variability (Winitz *et al.*, 1970). Pauling thus had to publish a correction to his highly cited remarkable paper.

In contrast to Pauling, PLIEM, the laboratory where I performed my GC–MS analyses, employs scientists who are experts in the field of metabolism. PLIEM has a rather long history in the identification and quantification of metabolites and the interpretation of metabolic data profiles. This experience is evident from their numerous valuable contributions (indicating the practical applications of the proper biological interpretation of data), ranging from screening to prenatal and postnatal diagnosis of IEMs, as well as their treatments. Highlights of these contributions include:

- One of the first GC–MS-based diagnoses made through PLIEM was for late-onset, biotin responsive, multiple carboxylase deficiency (Erasmus *et al.*, 1985), with the patient having a different phenotypic characteristic from the first case, described two years earlier. This contribution illustrates the importance of precise metabolite identification and quantification in the individual case — here, high levels of 3-hydroxyisovaleric acid and lactic acid (17.2 and 8.4 mmol/g creatinine, respectively), low levels of methylcitric acid (0.7 mmol/g creatinine), trace amounts of 3-hydroxypropionic acid and 3-methylcrotonylglycine, and levels of 3-methylcrotonic acid below the detection limit. The metabolites in low and undetected concentrations observed in the patient were identified because they were part of the metabolite profile of the index case of 1983. This humble experience illustrated two highly important points: (1) metabolic variability occurs even in individuals with the same inherited monogenetic deficiency; and (2) meticulous analytical characterisation of metabolites, linked to founded metabolic profiles, is vital for the identification of biomarkers. Dramatic clinical improvement could be seen within six weeks following treatment with biotin — from a relentless downward clinical profile of a hypotonic, unresponsive patient

with alopecia, hepatosplegomegaly, and respiratory complications to an infant that could sit unsupported, had head control, was cleared of respiratory infections, and became socially responsive.

- The second illustration from PLIEM is on the prenatal testing/screening for generalized glutathione synthetase deficiency in a family affected by this IEM (Erasmus *et al.*, 1993). The first child born into this family died of the disorder at the age of 6 weeks. Prenatal metabolic profiling and biomarker (3-oxoproline) identification, measured by stable-isotope dilution, was performed in two subsequent pregnancies on amniotic fluid. The results indicated that the fetuses were unaffected by the disorder. Therapeutic abortion was not recommended, and a healthy girl and boy were born from the second and third pregnancies, respectively. Neonatal urine metabolite analyses, as well as glutathione synthetase activity in haemolysates, were determined, and the results confirmed that both the infants were unaffected, and apparently had two normal alleles for the enzyme.
- The last example deals with isovaleric acidemia (IVA), an IEM known to occur rarely in various demographic locations, but one of the most common organic acidemias found in South Africa. Since 1983, a significant number of IVA cases have been identified in the approximately 20,000 Caucasian patients screened at PLIEM for metabolic defects. For her Ph.D. study, Marli Dercksen performed metabolic profiling of ten IVA patients, as well as phenotypic and genotypic investigations of the ten patients and three of their family members (Dercksen *et al.*, 2012). All patients were found to be homozygous for a single c.367 G > A (p.G123R) mutation, suggesting a common founder effect in the South African Causation population. Despite the genetic homogeneity of this IVA group, the clinical presentation varied widely, ranging from severe mental handicap and multiple episodes of metabolic derangement, to an asymptomatic state. This once again indicates the importance of not disregarding inter-individual metabolic variability in the interpretation of results from metabolomics investigations.

The upside of working with PLIEM, in analysing and interpreting the GC–MS-generated data and results, was that their knowledge of metabolism allowed for the ability to construct global metabolite profiles from the data, which provided me with an enormous advantage for my metabolomics study. The downside was that analytical-chemical, other technological, and data processing alternatives, which are key aspects in metabolomics research, are of lesser importance to PLIEM, and were therefore not explored. That being said, quality control, through their participation in various ERNDIM quality assurance schemes, is an ongoing, high priority for PLIEM (Figure 5–2), which was an additional advantage for me. Taken together, through the close link with PLIEM, the comprehensive metabolic interrelationships between various metabolites and metabolic pathways associated with acute alcohol consumption could be

uncovered, which paved the way for gaining new biochemical insights into the effects of this social practice.



Certificate of Participation and Performance for 2015

This certificate is to confirm that:

NORTH WEST UNIVERSITY POTCHEFSTROOM RSA, POTCHEFSTROOM (ERN0364)

participated in the following External Quality Assurance schemes organised by ERNDIM in 2015 and has achieved performance in the individual schemes as shown.

It is essential that the laboratory's own detailed annual reports for each scheme be referred to alongside this Certificate. For the Quantitative schemes these are available from www.erndimqa.nl and for the Qualitative schemes these are available from cscq.hcuge.ch/cscq/ERNDIM/ or by contacting ERNDIM at admin@erndim.org.

Quantitative Schemes

	Subscribed	Number of Analytes		
		Included in the scheme	For which results submitted by lab	Satisfactory Performance
Amino Acids (serum) (NORTHWEST)	Yes	28	21	12
Cystine in White Blood Cells ()	No	-	-	-
Lysosomal Enzymes (fibroblasts) ()	No	-	-	-
Purines and Pyrimidines (urine) ()	No	-	-	-
Quantitative Organic acids (urine) ()	No	-	-	-
Special Assays Serum (2015.01)	Yes	24	10	9
Special Assays Urine ()	No	-	-	-

Qualitative Schemes

	Subscribed	Submitted Results	Satisfactory Performance
Congenital Disorders of Glycosylation (serum)	No	-	-
Diagnostic Proficiency Testing (urine)	Yes	Yes	Yes
Qualitative Acyl Carnitines (DBS)	No	-	-
Qualitative Organic Acids (urine)	Yes	Yes	Yes
Urine Mucopolysaccharides (urine)	Yes	Yes	Yes

Signed on behalf of ERNDIM, 20-06-2016



Dr G Ruijter
Chair, ERNDIM Executive Committee

ERNDIM (European Research Network for evaluation and improvement of screening, Diagnosis and treatment of Inherited disorders of Metabolism) is a charitable foundation registered with the Maastricht chamber of Commerce (Registration number: 41078752) and coordinated by the ERNDIM Administration Office based in the Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester, UK, www.erndim.org

Figure 5-2: ERNDIM certificate of participation and performance of the PLIEM laboratory of North-West University.

5.4 Experimental and measurement design for an interventional GC–MS study

5.4.1 Background

From section 5.2 it is apparent that designing and executing a robust and carefully controlled standardized analytical and experimental protocol is of critical importance to address the potential problems and sources of variation as well as to ensure that reliable results and useful biological information will be obtained. The experimental design and experimental procedure — in terms of case selection, interventions and sampling — followed for the GC–MS analysis part of this intervention study is the same as discussed for the NMR analysis (Chapter 4). All the elements of the experimental design for the GC–MS analysis are illustrated and described in Figure 5–3, an adapted version of Figure 1 of Irwin *et al.*, 2016 (section 4.4.2).

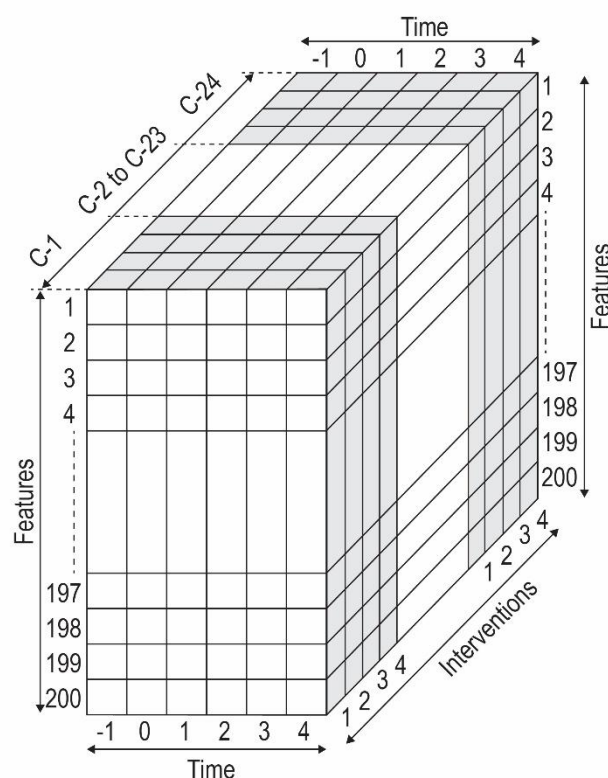


Figure 5-3: Representation of all elements of the experimental design. The tensor for the complete design consisted of four dimensions: the experimental subjects (total 24 – the shaded areas in the figure are for subjects 1 and 24), time points (6, hourly), interventions (4, including the “vehicle only”), and identified features (estimated 200, prior to analytical data pre-treatment). The total number of 115 200 data points obtained ($= 24 \times 6 \times 4 \times 200$) would then require bioinformatics analysis to uncover the information from the four interventions.

However, the measurement design of the study was modified for the GC–MS analysis, and is described comprehensively here, together with the organic acid extraction and derivatization, GC–MS analysis, and variable identification procedures. A good measurement design is important to define, control, reduce and correct analytical or instrumental variability — as well as other sources of variation — and also to ensure that the biological questions can be answered efficiently, taking into account measurement restrictions such as maximum batch size, and the number of measurement replicates and QC samples (Hendriks *et al.*, 2011). By using the PLIEM SOP for organic acid extraction, derivatization and analysis, and with the addition of internal standards and QC samples to the analytical runs, the modified measurement design presented here is aimed at assessing the effect of, accounting for and minimising the sources of variation referred to above.

5.4.2 GC–MS measurement design

Classically, organic acid profiling involves extraction with ethyl acetate and diethyl ether and trimethylsilyl (TMS) derivatization, followed by GC–MS separation and identification of the extracted features. This procedure is highly applicable to metabolomics studies using urine as biofluid for analysis, and is the basis for the method used in my study.

This study design produced a very large number of samples (more than 1000), and their practical analysis spanned a considerable time. This necessitated splitting the samples across multiple analysis blocks or batches. This is known to be problematic for GC–MS studies, since these instruments are prone to show drift within and between batches (Hendriks *et al.*, 2011). Therefore, repeatability and reproducibility are extremely important prerequisites in large-scale experiments, and in this case proved to be satisfactory through application of the KEMREP (kernel method for the assessment of repeatability and reproducibility) procedure for GC–MS analysis (Mason *et al.*, 2014), and by all experimental work prior to GC–MS analysis done by the same analyst.

A schematic representation of the measurement design, indicating the final batch outlay (consisting of experimental, QC and repeat samples per individual participant as well as hexane blanks), is shown in Figure 5–4 and detailed below.

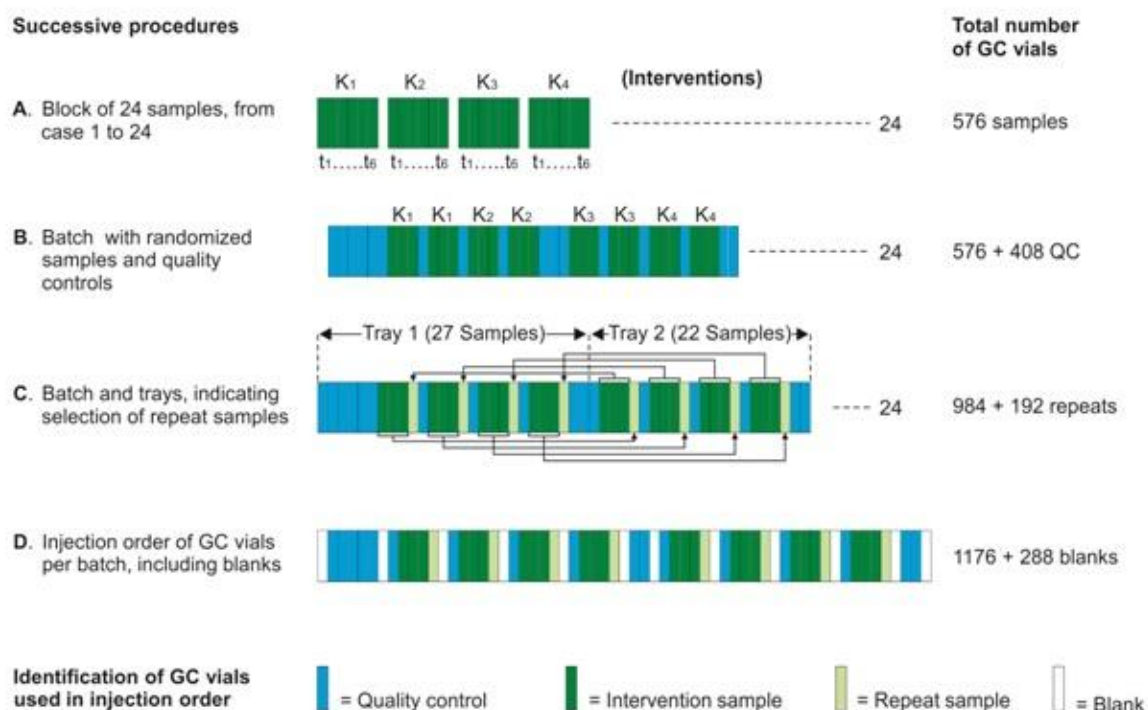


Figure 5-4: Measurement design for batches. Each of the 24 batches from the 24 experimental subjects consisted of the 24 experimental samples (6 x 4 for the time points and interventions), 17 QC samples, eight repeat samples and 11 hexane blanks. All QCs were identical, and the 24 experimental samples were randomly spaced within each batch as described. Every batch consisted of two trays, containing 27 and 22 samples (excluding blanks), respectively.

The 24 samples collected from each of the 24 subjects across the four interventions and the six time points per intervention formed the basis for the 24 batches to be analysed. Every batch was divided to consist of two trays, containing 27 and 22 samples (excluding hexane blanks), respectively. The 24 experimental samples were analysed per intervention, with the six samples from each intervention randomly selected. This means the six “vehicle only” samples were randomly placed first in tray 1, followed by the six “alcohol plus vehicle” samples, and the six “vehicle plus NAD” samples were randomly placed first in tray 2, followed by the six “vehicle plus alcohol and NAD” samples. The six samples from each intervention were separated into two sub-sections of three by a QC and a repeat sample.

In untargeted metabolic profiling studies, where all the metabolites in a sample are of interest, as well as in semi-targeted studies (such as the one described here), where all the metabolites of a specific class (in this case organic acids) are of interest, it is crucial to assess analytical repeatability and reproducibility (t'Kindt *et al.*, 2009). The measurement design therefore included the use of a large number of identical pooled QC samples to estimate, assess, and

possibly correct for, any batch effect (which would influence the within- and between-batch repeatability), or other interfering analytical aspects (Hendriks *et al.*, 2011) during the course of the whole analysis. In general, the greater the number of QC samples used during an analytical run, the higher the quality of the product matrix with regard to the reproducibility of low-intensity metabolites. However, the use of a larger number of QC samples significantly increases the total analysis time, which could increase experimental drift. It is therefore necessary to reach a compromise between these two factors (Kamleh *et al.*, 2012). During this investigation, each batch contained 17 QC samples in total, injected at regular intervals throughout the analytical batches: five at the start of the analytical run, two at the end of each tray, and one before each sub-section of three experimental samples. Each of the 17 QCs in each batch was prepared and derivatized individually, to provide 17 individual QC samples for each batch for GC–MS analysis. The QC samples can therefore be regarded as end-to-end QC samples, representing a profile for the analytical process as a whole — from sample preparation and extraction to GC–MS analysis. The data from the five QC samples preceding tray 1 (QCs 1–5) were used to condition the column and were excluded from the analysis of the interventions and in the determinations of repeatability and reproducibility as recommended (Gika *et al.*, 2007; Gika *et al.*, 2008; Zelena *et al.*, 2009).

Eight repeat samples were included in each batch, four in each tray, one placed directly after each sub-section of three experimental samples. The repeat samples included in tray 1 were experimental samples in tray 2 and vice versa. These repeat measurements were included to determine their repeatability between the two trays. Each batch also included 11 hexane blanks for analytical and demarcation purposes — hexane blanks are run through the GC–MS column from time to time through the course of a batch analysis to clean the column of compounds that become attached to it, and to prevent late-eluting compounds from previous runs to be measured in subsequent runs. These are common sources of variation mentioned in Table 5–3 that can be prevented with the use of blanks. Ideally, a blank should be run after each experimental sample, but this would increase the total batch analysis time too much in cases such as this where there is already such a large number of samples in each batch. Blanks can also be used to separate sections of samples in long analytical runs.

This final, mutually agreed upon, measurement design, presented here and meticulously followed for all the GC–MS analyses, was the outcome of lengthy planning sessions between, and inputs from, relevant experts: (1) The GC–MS-specialist advised on the use of QC samples at the beginning of each batch to condition the GC column, and on the addition of blanks. (2) The distribution of the QC samples within the batches was proposed by the statisticians to be optimal for assessing and possibly correcting for any within or between batch/tray effects. The addition of repeat samples was proposed since they could potentially be used, in addition to the

QC samples, for further analytical analysis on repeatability, as well as to assess the efficacy of any batch corrections applied. (3) The decision to analyse the experimental samples per intervention was based on a combination of biological and analytical arguments, although randomization of the six samples from each intervention was a precondition for the measurement design.

Extraction and derivatization for the samples of tray 1 was done on day one, after which the samples were placed in the GC autosampler and the GC–MS analysis was started. The GC–MS analyses of the samples from tray 1 took approximately 24 hours to complete. Therefore, since it is recommended that samples are derivatized as shortly before analysis as possible, the samples from tray 2 were prepared and derivatized only on the subsequent day, and placed in the GC autosampler directly after the samples from tray 1. The GC–MS analysis of each batch thus remained a continuous process, spanning a total of approximately 43.5 hours (24 hours for tray 1 and 19.5 hours for tray 2).

5.4.3 Extraction and derivatization of organic acids

Sample preparation vitally contributes to defining the array of metabolite classes covered. Any sample preparation protocol in a metabolomics perspective is a compromise between complete recovery of (ultimately) all compound classes, avoiding chemical or physical breakdown of labile metabolites and enzyme-mediated metabolite conversions, and producing a sample compatible with the separation technique to be used (t'Kindt *et al.*, 2009). In the case of GC–MS, the first step in sample preparation is the extraction of relevant metabolites, taking the class of metabolites and type of biological sample used into account. The extraction process is designed to: (1) isolate the maximum quantity of targeted/relevant metabolites; (2) remove interfering substances and reduce matrix effects; (3) concentrate trace metabolites; and (4) make the extract compatible with subsequent analytical procedures (such as derivatization) (Álvarez-Sánchez *et al.*, 2010).

GC–MS is used in the analysis of multiple classes of metabolites, but is biased against non-volatile and high molecular weight polar molecules. Analytes analysed with GC–MS therefore preferably need to be volatile, non-polar and thermally stable. However, many metabolites in urine are polar, ionic and relatively non-volatile, and therefore the second step in sample preparation involves chemical derivatization of the metabolites prior to analysis (Want *et al.*, 2010). Derivatization refers to multiple, non-spontaneous, parallel, competing reactions of metabolites from various chemical classes with the same reagent. Various forms of chemical derivatization of metabolites have been proposed and described (Halket *et al.*, 2005), of which silylation — where active hydrogen atoms in polar functional groups are replaced with TMS groups, resulting in decreased polarity and increased volatility and thermal stability — is the

most commonly used in metabolomics studies, since it covers a wide range of metabolites. Silylation was therefore also selected as the method of derivatization for this study. Derivatization using silylation is an important step, but can also introduce various potential sources of variation to the metabolomics data (see Table 5–2).

The amount of urine sample, internal standard (IS) and derivatization reagents used for GC–MS analysis is based on the creatinine value of the urine. Adjustments of volumes during sample preparation according to creatinine accounts for any concentration variances between urine samples — that is, more reagents and less urine are used for analysis if the urine sample is concentrated and vice versa for diluted urine samples). A 1 mL aliquot of each urine sample was used for creatinine determination, performed by an external pathology service.

For their successful participation in the ERNDIM qualitative measurement of organic acids in urine (Figure 5–2), PLIEM use an SOP for organic acid extraction and derivatization, followed by GC–MS analysis. This is also the method most often applied to metabolomics investigations performed in this laboratory (Dercksen *et al.*, 2013; Reinecke *et al.*, 2012). Based on this, this method is regarded as satisfactory with regard to repeatability and reproducibility, and was therefore also the method used in the present GC–MS investigation. Details of this method, as well as details on the GC–MS analysis and variable identification procedures, are given in the Supplementary Information of Irwin *et al.*, 2018 (section 5.7).

5.5 Metabolite identification and classification

Data generation and analytical peak assignment of the samples, analysed in different batches (see Figure 5–4), produced a metabolomics matrix, including data from all experimental, QC and repeat samples for all times and all interventions. For practical reasons, the 100 QCs with the maximum value for the mean concentrations of all features were selected from the total number of QCs analysed, and used for the biological description, identification and classification of each of the features present in these representative samples. QC samples were used for this purpose based on the underlying assumption that all metabolites present in the experimental samples should theoretically be observable in these samples.

A total of approximately 550 targets were detectable in the 100 QC samples used following GC–MS analysis. From these, 176 variables were identified using the AMDIS (Automated Mass spectral Deconvolution and Identification System) software and both the PLIEM library and the commercially available NIST (National Institute of Standards and Technology) database (NIST 17 Main EI MS Library). Data from all other targets were discarded, since they were not consistently above the detection limit or could not be unequivocally identified. The 176 identified variables included two internal standards and two features still labelled as “Unknown” in the

libraries of the databases used. These two “Unknowns” were later identified as two artifacts commonly found in GC–MS analyses. The data from these four features were removed from subsequent analyses. All data of the remaining 172 identified variables were normalized relative to the data from the internal standard, 4-phenylbutyric acid. All variables identified above the detection limit of the equipment used were quantified relative to the concentration of the internal standard, and expressed as relative concentrations in $\mu\text{mol}/\text{mmol}$ creatinine. Several of the variables detected were present in concentrations just above the detection limit, and were not observed in all QC samples.

The Human Metabolome Database (www.hmdb.ca) was used as the reference for the biological description and classification of each of the remaining 172 variables. Information on variables not included in the HMDB was obtained from other established chemical databases or from the literature, and failure with this resulted in classification of 7 variables as “no annotation”. Based on this information, I propose a classification of the 172 identified variables according to the following major categories:

1. Endogenous substances

Substances that originate from within a living organism, tissue, or cell, and are produced naturally in the body, usually by metabolic processes (Dictionary.com Dictionary, 2017; Oxford Reference, 2017).

1.1 Primary and secondary human metabolites

Low molecular weight chemical compounds, found in humans, that are the intermediates and end-products of normal primary or secondary metabolism, or are produced as a result of an endogenous (an IEM) or exogenous (an intervention) perturbation. Primary metabolites, including amino acids, alcohols, vitamins, organic acids and nucleotides, are indispensable for growth, development and reproduction of an organism. Secondary metabolites are typically organic compounds produced through the modification of primary metabolites, are not usually required for primary metabolic processes, but may have important ecological and other functions (Berdy, 2005; Harris, 2014; Lumen Learning, s.a.; Meštrović, 2015).

1.2 Biotransformation products

Any substances produced in the body through biotransformation processes — the chemical modifications (usually through enzymatic actions) made by an organism on foreign chemical compounds in the body, such as drugs or other toxic substances — generally referred to as xenobiotics — in order to make them less toxic and more water-soluble (and therefore more easily excretable from the body) (Kebamo *et al.*, 2015; Macherey & Dansette, 2008; Free Dictionary Medical Dictionary, 2003).

1.3 Metabolites originating from the gut microbiome

Intermediates and/or end-products of the metabolism of dietary constituents by the bacteria of the human gut microbiota, including various short-chain fatty acids — produced by microbial digestion/fermentation of indigestible dietary fibre — and certain vitamins (Lin & Zhang, 2017; Richards *et al.*, 2016; Smith *et al.*, 2013; Thorburn *et al.*, 2014).

2. Exogenous substances

Substances that are present (and possibly active) in an individual organism or cell, but which originated from outside the organism, such as food, drugs, pathogens, etc. (Merriam-Webster Dictionary, 2017).

2.1 Substances related to the diet

Any ingested dietary substances, or substances or metabolites produced through the metabolism of anything that was ingested by an individual. This includes metabolites related to food, drink and drug metabolism.

2.2 Artifacts

Substances observed in scientific investigations or experiments that are not naturally or normally present and do not reflect the original sample or specimen, but occur as a result of the preparative or investigative procedure (mostly due to the derivatization process) and/or the equipment used (Oxford Living Dictionaries, 2017). These include the products of unwanted or unexpected chemical reactions that can occur during experimental analysis (Biology Online Dictionary, 2005).

2.3 Substances from the environment

Chemical reagents or toxic agents that an individual is exposed to, or ingests, that are eliminated from the body.

A table giving the details of each of the 172 identified and classified variables is reproduced in Table S2 of the Supplementary Information of Irwin *et al.*, 2018.

5.6 Additional experiences with sources of variation

The SOP used during this investigation for the extraction and derivatization process is standardized to optimally deal with the analytical sources of variation mentioned in section 5.2. However, despite following this frequently used and reliable SOP, careful sample handling and analysis, and consideration of the quality control procedures to minimize these problems, many additional sources of variation were encountered, as discussed in the following section. Some of

these could be accounted for or overcome, whereas others rendered some of the data unusable.

5.6.1 Artifact formation

Artifacts are a common problem in analytical chemistry and are frequently encountered in a variety of chromatography techniques (Little, 1999). It is well known that artifacts result from various sources — mostly related to incorrect or unfavourable derivatization conditions, such as incomplete derivatization or derivatized samples being stored at room temperature for long periods of time (Dunn & Ellis, 2005). In an in-depth study on artifact formation (Little, 1999), Little reveals some of the sources of artifacts, including that: (1) certain functional groups (such as aldehydes, ketones, carboxylic acids, esters, amides and phenols under certain conditions) are more likely to form additional unexpected derivatives from silylation reagents and their by-products; and (2) derivatization reagents can react with themselves, other organic reagents, organic solvents, or inorganic reagents to form various inorganic reagent artifacts, solvent artifacts, and silylation reagent artifacts.

In addition to artifact formation, other problems/sources of variation associated with derivatization include the formation of unexpected derivatives and metabolite conversions, the formation of multiple derivatives and multiple peaks for a single compound in a GC–MS spectrum (leading to peak overlapping), and the degradation of some low-stability TMS derivatives (especially amino acids). These factors complicate the interpretation of the spectra, making sample compound identification, determining the number of compounds in a sample, and calculating the correct concentrations of compounds difficult (Halket *et al.*, 2005; Little, 1999; Quéro *et al.*, 2014). The sources of variation associated with derivatization, as well as the quality control procedures that can be implemented to minimize these problem, are discussed fully in Table 5–2.

During this investigation, numerous artifacts were formed and identified following GC–MS analysis — indeed, approximately 10% of the 172 variables identified were classified as artifacts. It was noticed, for example, that several artifacts were formed through intra- or intermolecular reactions or rearrangements. Some of these artifacts are well known and described in the literature — for example, the formation of lactyl-lactic acid. However, this study also revealed various novel artifacts, which we are considering reporting elsewhere (in *Analytical and Bioanalytical Chemistry*). Further details on the artifacts are therefore not described here.

An analytical complication that contributed to the large occurrence of artifact formation was that this study required the analysis of an extremely large number of samples, which spanned a

considerable time. It is recommended, as standard practice, that samples are derivatized as soon as possible before GC–MS analysis. Therefore, in order to deal most effectively with the large number of samples, and to minimize the time that samples would be derivatized before analysis, samples were split into multiple batches (each batch containing the samples of one participant) and each batch was split into two trays (as discussed in section 5.4). Extraction and derivatization of the samples of the first tray of a batch was done on one day, after which the samples were placed in the GC autosampler and the GC–MS analysis was started. The samples from the second tray of the batch were prepared and derivatized only on the subsequent day, and placed in the GC autosampler directly after the samples from the first tray. However, since GC–MS analysis of each sample takes approximately 50 minutes, even splitting each batch (containing 49 samples) into two separate trays meant that the total GC–MS analysis times of the first and second trays (containing 27 and 22 samples, respectively) were still approximately 24 hours and 19.5 hours, respectively.

The QC samples included in each batch were also used to monitor the formation of artifacts. When the results from the QC samples were compared, it was evident that a considerable number of especially the novel artifacts were only observed in the samples analysed nearer to the end of each of the trays, which suggests that their formation resulted from the long analysis time of each tray. Owing to these long total analysis times, the last QC sample of the first tray of each batch was analysed only approximately 23 hours after the first one, meaning that these last samples were left at room temperature in the autosampler for 23 hours after being derivatized. In addition, these samples had a much longer time for the derivatization process to take place — even though it occurred at room temperature — leading to the formation of many more and different derivatives and artifacts, including artifacts that would not have formed if the samples were analysed more rapidly after being derivatized. The last QC sample of the second tray of each batch was analysed approximately 18.5 hours after the first QC sample of those trays, creating similar problems and artifact formation. This observation clearly illustrated that, owing to the long analysis times of each tray, derivatizing such large batches in only two separate instances was not sufficient to prevent excessive artifact formation. However, we assumed that the variables identified as artifacts could be eliminated from the data set prior to statistical analysis, and therefore would not further adversely influence the quality and interpretation of the results obtained from the data of the biologically relevant metabolites. This assumption is somewhat supported by the observation that the presence or absence of the high levels of hippuric acid, following the interventions, did not influence the outcome of the statistical analyses of the important variables related to the alcohol consumption. This finding is detailed in the Supplementary Information of Irwin *et al.*, 2018.

5.6.2 Repeatability of QC sample analyses

As mentioned in section 5.4, QC samples were included in all of the batches to assess analytical repeatability. Simply stated, if repeatability was absolute, all of the QC samples analysed would give identical GC–MS profiles. In practice, however, this cannot be fulfilled, and even maintaining a high level of repeatability is often difficult, since there are many factors and different potential sources of variation involved in metabolomics investigations that might compromise run-to-run repeatability (see section 5.2) (Fukusaki & Kobayashi, 2005; Gika *et al.*, 2008). In a satisfactory analysis, the QC samples are expected to cluster closely together and show little to no variation or time-related trends when performing a statistical technique such as PCA (Gika *et al.*, 2007; Gika *et al.*, 2008; Pasikanti *et al.*, 2008). This is an indication of the high repeatability of the analytical run, as well as the reliability of the data obtained from the rest of the analyses in that run. Only if this is the case can differences between different test samples be attributed to biological variation as opposed to analytical variation (Gika *et al.*, 2007; Gika *et al.*, 2008). QC samples showing high variability — which would indicate sample instability under the conditions in which the samples were maintained during analysis — should be easily detectable from PCA data (Gika *et al.*, 2007; Kamleh *et al.*, 2012). Evidence of such outlier QC samples should lead to questions regarding the reliability of the rest of the data from that analytical run.

Once the experimental GC–MS analyses of all the batches had been completed, the spectra generated from the analyses of the QC samples were compared. This comparison showed large between-batch, within-batch and even within-tray variations. This variation made it very difficult to justify using these samples for any type of batch correction, and, subsequently, no such correction was done. The comparison of the QC samples also indicated that, in general, the repeatability of the analyses of the second tray of a batch (“vehicle plus NAD” and “vehicle plus alcohol and NAD” interventions) was much less than that of the first tray of the same batch (“vehicle only” and “alcohol plus vehicle” interventions). Owing to this observation, the results from all the analyses of the samples in tray 2 were regarded as unreliable and therefore not used for further data analyses.

Given the logistical realities of the metabolomics laboratory used, the data from each batch of analyses could not be analysed separately per batch, and were instead stored electronically in the GC–MS database until all the analyses from all 24 batches were completed. Only then were all the data from all 24 batches analysed simultaneously. This, however, meant that the problems regarding the repeatability of the QC samples within and between the batches, as well as the lower repeatability of the QC samples in tray 2, only became known after all the analyses had been completed. Since nothing could be done at this stage to obtain better, more repeatable results, except for repeating the entire sample analysis process — which was not an

option due to time constraints — the data set could not be used to assess the applicability of the measurement design, and it was decided to discard the idea of using the QC and repeat samples for any type of batch correction. Also, the decision was made not to use the data from the samples in tray 2 for any further data analyses.

It became evident, however, that the large observed variations were most probably due to the logistical and instrumental aspects of the GC–MS analytical platform and the problems associated with it that were encountered during this study (section 5.6.3 below) — comparison of the same pooled QC samples, but analysed with NMR, yielded much more repeatable results and excellent CV% values (section 4.3 and Figure 4–2). Additionally, PCA was performed on the QC samples analysed with NMR, and the results could successfully be used for outlier detection (section 3.1 of the Supplementary Information of Irwin *et al.*, 2016). These observations from my data indicated the occurrence of higher repeatability in the NMR analysis compared to the GC–MS analysis, which, together with the amended approach followed for analysing the GC–MS data (discussed in section 5.6.4 below), proved the value of using both of these complementary analytical platforms in the present investigation.

5.6.3 Laboratory logistics

In metabolomics studies, analytical repeatability is a prerequisite for extracting reliable information from the data generated. The general aim of metabolomics studies is the identification of mostly subtle differences in biological systems under defined conditions. Following standardized protocols for sample handling, storage and analysis is vital to maintain the constant measurement conditions necessary to identify these subtle biological differences. The smallest changes in the sample treatment or analysis can introduce non-biological variation, which complicates data interpretation. Therefore, all sample analyses should ideally be performed by a single analyst, using the same analytical method and instrumentation, and under identical experimental and environmental conditions. In order for the metabolomics data obtained to be considered reliable and repeatable, the variation between the data should fall within a defined bracket of deviation (that is, an acceptable critical value for measurement of repeatability) (Fukusaki & Kobayashi, 2005; Mason *et al.*, 2014; Rist *et al.*, 2013).

During this study, various obstacles were encountered which complicated the prerequisite factors for repeatability stated above. The large number of samples collected during this study were all analysed by the same analyst (myself), and using the same analytical method (the PLIEM SOP for extraction and derivatization of urinary organic acids), as is recommended for good repeatability. However, the experimental and environmental conditions of the analyses could not be guaranteed to be identical — during the time of my GC–MS sample analysis, the existing PLIEM laboratory had to be demolished to make room for a new building for the TIA

Metabolomics Platform that had to be erected next to the Biochemistry building, The PLIEM laboratory, and all its equipment, was therefore temporarily transferred to another laboratory in the Biochemistry building. The analyses of my samples were therefore partly performed in the old, temporary and new PLIEM laboratories, which inevitably resulted in changes in the environmental conditions during sample analysis. Working in the temporary laboratory also created several logistical problems, including time constraints on availability of the equipment, and the need to share the GC–MS apparatus. This meant that my samples could not always be analysed when they ideally needed to be. Because of this, and in view of the large number of samples that had to be analysed, the total analysis of all the samples (more than 1000 including QCs and repeats) spanned several months, which resulted in additional seasonal temperature variation during sample analysis. These factors contributed to the noticeable variation in the repeatability of the QC samples. However, these complications also provided the opportunity for “lessons learned” (see Chapter 6), which remains a key component of education, even at the advanced level of a Ph.D. study.

5.6.4 Problem solving

As mentioned above, the data generated from the samples in the second tray of each batch posed repeatability problems that could not be solved through batch corrections. As a consequence, these data could not be used and no metabolomics analysis could therefore be performed on the effect of NAD supplementation from the GC–MS data. This was fortunately not the case, however, for the analyses of the samples from the first tray of most of the batches. The same high analytical precision method that was used to identify the 172 variables in the QC samples (detailed in section 5.5) was used to identify and quantify these same variables in data from all tray 1 samples (that is, the “vehicle only” and “alcohol plus vehicle” interventions) from 12 of the batches. The PCA scores plot in Figure 5–5 illustrates the variation in the QC and experimental samples from these quantified batches. The variation in the QC samples is much less than the variation in the experimental samples, which indicates the repeatability of the QC samples. This result gives an indication of the sufficient analytical quality of the experimental data obtained from these batches.

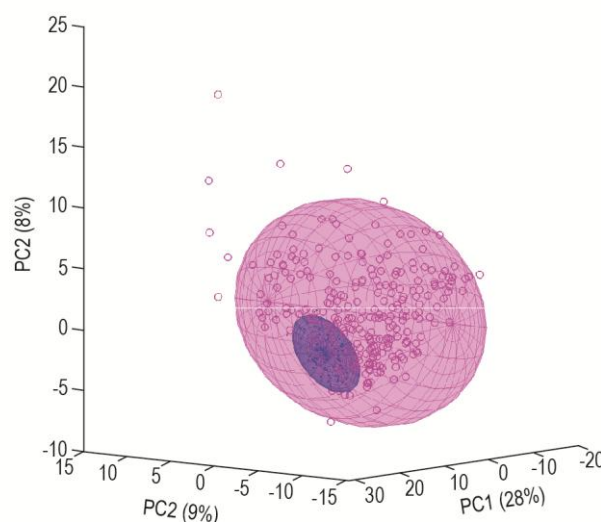


Figure 5-5: PCA scores plot of the QC samples (blue) and experimental samples (pink) from the 12 batches selected to be quantified. The close clustering of the QC samples compared to the experimental samples indicates that the variation in the QC samples is much less than the variation in the experimental samples, which suggests that the experimental data obtained from these batches are of suitable analytical quality to be used to investigate the effect of acute alcohol consumption.

This method of meticulously quantifying the metabolites in each sample individually is highly reliable, but also extremely tedious and time-consuming, and therefore, given time constraints, only the tray 1 samples of 12 of the 24 batches (thus, data from 12 experimental subjects) were analysed in this way. From the 172 variables, 120 were found to have biological relevance (see Table S2 of the Supplementary Information of Irwin *et al.*, 2018), and were used for the metabolomics analysis of the effect of alcohol consumption. The outcomes of this data analysis proved to be valuable, and produced some novel insights. These findings are presented in section 5.7 as a peer-reviewed paper (Irwin *et al.*, 2018).

5.7 Outcomes of a GC–MS metabolomics study on acute alcohol consumption (Irwin *et al.*, 2018)

SCIENTIFIC REPORTS

OPEN

GC–MS-based urinary organic acid profiling reveals multiple dysregulated metabolic pathways following experimental acute alcohol consumption

Received: 29 November 2017

Accepted: 22 March 2018

Published online: 10 April 2018

Cindy Irwin¹, Lodewyk J. Mienie¹, Ron A. Wevers², Shayne Mason¹, Johan A. Westerhuis^{1,3}, Mari van Reenen⁴ & Carolus J. Reinecke¹

Metabolomics studies of diseases associated with chronic alcohol consumption provide compelling evidence of several perturbed metabolic pathways. Moreover, the holistic approach of such studies gives insights into the pathophysiological risk factors associated with chronic alcohol-induced disability, morbidity and mortality. Here, we report on a GC–MS-based organic acid profiling study on acute alcohol consumption. Our investigation — involving 12 healthy, moderate-drinking young men — simulated a single binge drinking event, and indicated its metabolic consequences. We generated time-dependent data that predicted the metabolic pathophysiology of the alcohol intervention. Multivariate statistical modelling was applied to the longitudinal data of 120 biologically relevant organic acids, of which 13 provided statistical evidence of the alcohol effect. The known alcohol-induced increased NADH:NAD⁺ ratio in the cytosol of hepatocytes contributed to the global dysregulation of several metabolic reactions of glycolysis, ketogenesis, the Krebs cycle and gluconeogenesis. The significant presence of 2-hydroxyisobutyric acid supports the emerging paradigm that this compound is an important endogenous metabolite. Its metabolic origin remains elusive, but recent evidence indicated 2-hydroxyisobutyrylation as a novel regulatory modifier of histones. Metabolomics has thus opened an avenue for further research on the reprogramming of metabolic pathways and epigenetic networks in relation to the severe effects of alcohol consumption.

Notwithstanding the encyclopaedic information on alcoholism, the WHO asserts that alcohol remains one of the world's leading risk factors for disability, morbidity and mortality — 5.9% of all deaths worldwide are attributable to alcohol consumption, exceeding those from HIV/AIDS (2.8%), violence (0.9%) or tuberculosis (1.7%)¹. The tenth edition of the International Classification of Diseases lists at least 25 chronic conditions that are entirely attributable to alcohol; alcohol is also a risk factor in certain cancers, some tumours, numerous cardiovascular and digestive diseases, and many neuropsychiatric conditions². Brain image studies have revealed changes in brain structure during the progression from adolescence to adulthood^{3,4}, a critical period characterized by increased brain connectivity and maturation of brain neural circuits. These changes are highly susceptible to the effects of exogenous substances, which likely include alcohol^{5–7}, making children and adolescents especially vulnerable to alcohol-related harm⁸. It was reported from a recent survey that approximately 50% of children (aged 11–14 years) in the UK have consumed alcohol, and 33% of adolescents (15–16 years) admitted to having experienced at least one episode of acute alcohol intoxication in the month preceding the survey⁹.

¹Centre for Human Metabolomics, Faculty of Natural Sciences and Agriculture, North-West University (Potchefstroom Campus), Private Bag, X6001, Potchefstroom, South Africa. ²Radboud University Nijmegen Medical Centre, Translational Metabolic Laboratory, Department of Laboratory Medicine, PO Box 9101, 6500 HB, Nijmegen, The Netherlands. ³Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands. ⁴Department of Statistics, Faculty of Natural Sciences and Agriculture, North-West University (Potchefstroom Campus), Private Bag, X6001, Potchefstroom, South Africa. Correspondence and requests for materials should be addressed to C.J.R. (email: carools.reinecke@nwu.ac.za)

Human genome-wide association studies (GWAS) have identified polymorphisms and candidate genes associated with individuals' innate risks for alcohol dependence^{10–12}. With metabolomics involving large-scale molecular studies of metabolic systems, it provides novel and complementary approaches to GWAS studies^{13–15}. Additionally, metabolomics has revealed the extensive perturbations of various metabolic pathways in response to chronic alcohol consumption, which underpin alcohol-induced disease states. Such studies show much promise for disease profiling and biomarker identification of the conditions associated with and arising from chronic alcohol consumption¹⁶. The burden attributable to acute alcohol consumption (rapid ingestion of alcoholic beverages) or binge drinking (drinking too much too fast) has been shown to be similarly high^{17–19}. Impediments associated with acute and binge drinking include perturbed metabolism (e.g. glycogen depletion, acidosis and hypoglycaemia)²⁰, damage to intestinal epithelial cells²¹, neurobiological diseases²², and ultimately even abrupt premature death due to physical or mental disabilities^{23,24}.

Several important observations on acute alcohol consumption have recently been obtained from animal-based studies:

1. A rodent “intra-gastric feeding model” was investigated with ultra-high performance liquid chromatography–time of flight mass spectrometry (UHPLC–TOFMS) technology to determine changes in global metabolite profiles for plasma and urine from alcohol treated rats and mice compared to control animals²⁵. Apart from several other observations, the researchers reported changes in the concentrations of 5-hydroxytryptophan and xanthurenic acid, both of which are intermediates in tryptophan metabolism. These observations provide further insight into the association of liver metabolism in response to ethanol exposure.
2. A study on the immune response of healthy individuals (11 males and 14 females, aged 21–56 years) with no history of alcohol use disorder, revealed that acute binge drinking resulted in a rapid increase in serum Gram-negative bacterial endotoxin lipopolysaccharide (LPS) and bacterial 16S rDNA. LPS is a potent trigger of the inflammatory cascade via activation of the Toll-like receptor 4 (TLR4) complex and increase in the portal and/or systemic circulation in several types of chronic liver diseases²⁶. The elevation of bacterial 16S rDNA levels after acute binge drinking indicated transient gut-derived microbial translocation as a likely mechanism for the serum LPS increase. The authors suggest that the increased serum levels of bacterial products following acute consumption might contribute to innate immune responses and potentially to the behavioural effects associated with alcohol binge drinking. The study also draws attention to potential perturbations in metabolite profiles which may be due to the role of the gut and the microbiome in binge drinking. This point is substantiated by GC–MS urinary metabolomics results from another study, which revealed metabolite differences between Sprague–Dawley and Wistar rats following different perturbations, including consecutive acute ethanol interventions²⁷. These results directed to different metabolic pathways and differences in the intrinsic metabolism and symbiotic gut microflora between these animal strains.
3. A review of the analytical technologies used in profiling studies of animal or human serum, plasma, urine and tissue samples, obtained following exposure to alcohol, summarizes a range of endogenous metabolites that have been proposed as potential ethanol consumption-related biomarkers²⁸. This range of biomarkers provides biochemical insights that are essential for understanding the effects and mechanisms of ethanol toxicity.

Notwithstanding the potential of metabolic profiling, a recent review noted that few studies on acute alcohol consumption have been undertaken in humans²⁹, recognizing that these models could provide a basis for studying the biochemical effects of prolonged ethanol exposure, as well as to potentially identify biomarkers for monitoring the progression of alcoholism in man²⁵.

We thus postulated that our metabolomics approach could provide further insights into the metabolic signature which arises from a single excessive dose of alcohol. However, the response to experimental alcohol consumption is complex, as shown by the diverse observations noted from the study on two related animal strains²⁷. Alcohol studies in humans, likewise, vary greatly according to the extent and method of usage (chronic or acute), individual variability (genetic and behavioural), and environmental factors. Also, individuals' attitude to their consumption habits is a private issue, and important ethical considerations (such as restricted case participation and policy guidelines) have to be taken into account during the selection of sufficient and appropriate participants for the laboratory assessment of alcohol use³⁰. Therefore, recording the influence of alcohol in laboratory studies involving a select group of moderate-drinking young men, as we present here, is complex from several points of view.

Despite these limitations and qualifications, the application of metabolomics to intervention or challenge studies is a preferred practical approach towards a holistic understanding of the effects of consumed substances on metabolism^{31,32}. Intervention studies produce extensive data sets due to longitudinal (time-dependent), multi-subject (several experimental participants), multi-group (number of interventions), and multivariate data (numerous metabolic variables) inputs^{33,34}. Our previous intervention study on consuming commercial flavoured water with a benzoic acid preservative³⁵ indicated that time-dependent metabolomics investigations, using designed interventions, provide a way of interpreting the variation induced by the different factors of a designed experiment. This approach has the potential to significantly further our understanding of normal and pathophysiological perturbations of endogenous or exogenous origin. Here we report on a metabolomics study which used commercial flavoured water as a vehicle³⁵ for vodka consumption, to simulate acute but controlled alcohol consumption (resembling a single incident of acute or binge drinking) in young male participants. All metabolomics studies are inherently hampered by an analytical inadequacy to provide a comprehensive coverage of the metabolome. Given the known main pathways of alcohol metabolism, we selected a gas chromatographic–mass spectrometric (GC–MS) approach to analyse the

urinary samples collected during this time-dependent cross-over study. These analyses generated an extensive data set, from which (by using conventional and extended statistical methods) we could reveal a multitude of disturbances in the urinary organic acid profiles over time due to the alcohol intervention.

RESULTS

Effect of the acute alcohol intervention across all cases. Observations of the behaviour of the participants during the intervention stage indicated a moderate degree of acute alcohol consumption characteristic of the early phase of a binge-drinking event by young male students³⁶. To determine the effect of the alcohol intervention, the data sets of 120 organic acid metabolites, generated from the samples collected at times 1, 2, 3 and 4 hours following the intervention, were compared with the data set from the samples collected prior to the intervention (time 0), and presented for unsupervised principal component analysis (PCA) (Fig. 1a–d). Subsequently, a supervised partial least squares discriminant analysis (PLS–DA) was performed to maximize the discrimination between the controls (time 0) and the subsequent hourly-collected data. The PCA scores plots showed some differentiation 1 hour after the alcohol consumption (Fig. 1a), followed by complete separation after 2 hours, and a progressive return to the time 0 profile after 3 and 4 hours. The PLS–DA scores plots (Fig. 1e–h) showed a complete separation for all four times following alcohol consumption relative to time 0.

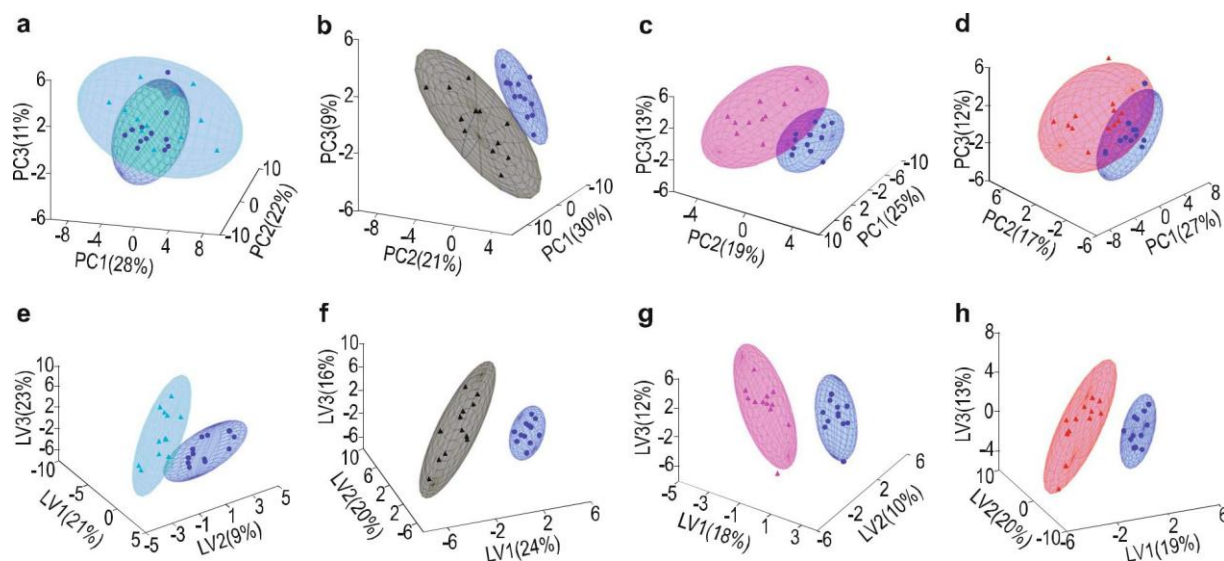


Figure 1. PCA and PLS–DA scores plots following alcohol consumption. Input data were from the 120 quantified metabolites for time 0 vs time 1 (a,e), time 0 vs time 2 (b,f), time 0 vs time 3 (c,g), and time 0 vs time 4 (d,h) following alcohol consumption. The samples for time 0 were collected just prior to the intervention, and are therefore regarded as the control samples. The figures were based on the non-paired method of analysis. Results from the paired method are shown in Supplementary Fig. S2.

For better insight into the time effect of the intervention, two complementary approaches were used, both of which are founded on Principal Component Analysis (PCA). The first, ANOVA Simultaneous Component Analysis (or ASCA), enabled us to partition variation similar to an analysis of variance (or ANOVA) approach using the experimental design, and summarize variation observed across many variables through projection to a lower dimensional space. The scores plot resulting from the ASCA model (Fig. 2a) indicates a marked change one hour after the intervention. After 3 hours a state of homeostasis is reached as indicated by the similarity between times 3 and 4. This is, however, a new state of homeostasis, since it is not comparable to the state prior to the intervention, as will be discussed below. Though time 2 appears to show a larger perturbation compared to time 1, following the intervention, time 2 also shows a progression towards the new state of homeostasis, making it less indicative of the initial alcohol effect. The ASCA model partitioned variation from various sources in the design (that is, time and participant) by averaging out individual effects. Secondly, performing PCA on the dataset unfolded in time provides another view on the variation observed, which ties repeated observations together into a single profile. The scores from such a model provide a complementary and, in a sense, confirmatory summary of the effect of the alcohol consumption in time, as indicated by the score centroids shown in Fig. 2b. It is evident that there was a change in the organic acid profile (i.e. across all cases) over time, with the most extensive change visible 1 hour (time 1) after the intervention. The centroids plotted for times 2 to 4 complement the interpretation made from the ASCA.

Based on the statistical criteria, a total of 13 variables (Table 1) were identified as those responsible for the separation between the control samples and those collected one hour after alcohol consumption. Shown in Table 1 are the VIP values, *p*-values and FC values at time 1 relative to time 0, as well as the mean and standard deviation values for all times. The FC and *p*-values for time 4 relative to time 0 are also included to show the eventual diminishing effect of the alcohol on the 13 metabolites. In addition, since ethanol data is essential in the context of this intervention study, these values are included for reference purposes. However, ethanol cannot be reliably detected with the GC–MS method used here, and therefore the quoted urinary ethanol values were measured through

a targeted proton nuclear magnetic resonance [$^1\text{H-NMR}$] spectroscopy analysis. The full metabolite profile that became apparent through this NMR metabolomics study will be reported in a subsequent paper.

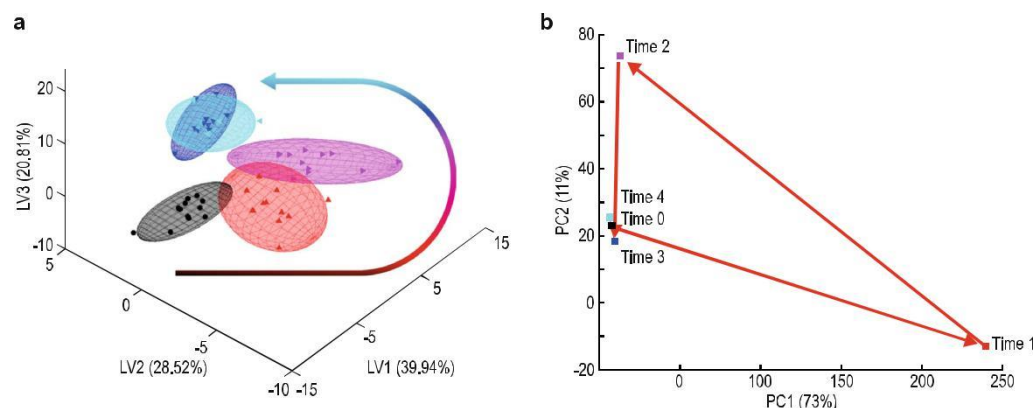


Figure 2. Multivariate approaches to indicate the time effect. **(a)** ASCA scores on the first three latent variables (LV1 to LV3), along with 95% confidence ellipsoids for the centroids, are shown and colour-coded according to time following the intervention (0, 1, 2, 3 and 4 hours shown in black, red, pink, dark blue and light blue, respectively), with the arrow showing the time-dependent trend, using the same discriminating colour sequence. **(b)** Unfolded PCA scores centroids for each time point based on the first two principle components (PC1 and PC2) are shown. The colours of the centroids are the same as for the ASCA and the direction of the trajectory is indicated by the red arrows, starting from time 0.

Metabolites ^a	PLS-DA VIP ≥ 1.0 (Controls vs 1 hr)	Summary statistics (Controls vs 1 hr)		Mean concentrations of metabolites [Std. Dev.] μmol metabolite/mmol Cr					Summary statistics (Controls vs 4 hr)		HMDB Ref. values μmol metabolite/ mmol Cr
		[WRT] ₀₋₁	[FC] ₀₋₁	t=0	t=1	t=2	t=3	t=4 hr			
Hippuric acid	19.6	0.002	+3.2	948 [737]	3037 [1357]	2775 [2051]	1157 [1111]	726 [764]	0.239	-1.3	27.9–932.7
Lactic acid	14.4	0.015	+27.4	8.73 [3.65]	239 [412]	28.1 [19.7]	15.5 [14.7]	12.1 [8.51]	0.209	+1.4	3.9–9.8
Fumaric acid	7.24	0.003	+6.4	0.78 [0.47]	4.96 [4.53]	5.88 [6.08]	2.62 [1.94]	2.8 [2.15]	0.006	+3.6	0.75–1.2
Vanillylmandelic acid	5.03	0.034	-1.6	15.4 [3.61]	9.87 [7.39]	2.96 [2.32]	1.87 [1.11]	1.75 [0.99]	0.002	-8.8	1.1–1.7
2-Hydroxybutyric acid	4.06	0.034	+4.3	2.68 [1.54]	11.4 [16.0]	3.85 [2.68]	2.26 [1.16]	2.52 [2.15]	0.695	-1.1	1.2–6.9
Succinic acid	3.88	0.010	+2.6	3.81 [3.81]	9.83 [7.90]	20.2 [24.6]	12.8 [12.8]	12.7 [12.3]	0.004	+3.3	4.9–14.9
3-Hydroxybutyric acid	2.76	0.041	+5.9	0.40 [0.54]	2.39 [2.96]	0.59 [0.55]	0.47 [0.40]	0.77 [0.68]	0.182	+1.9	1.4–2.2
2-Ethylhydracrylic acid	1.66	0.010	+2.1	1.09 [1.00]	2.30 [1.73]	2.53 [2.78]	1.76 [1.61]	1.86 [1.47]	0.015	+1.7	1.3–2.9
3-Hydroxyisobutyric acid	1.56	0.034	+1.9	3.01 [1.27]	5.76 [3.86]	3.31 [2.05]	2.74 [1.25]	3.44 [1.95]	0.239	+1.1	4.1–19.0
2-Hydroxyisobutyric acid	1.39	0.034	+1.8	11.1 [4.03]	20.5 [14.2]	20.5 [13.4]	12.8 [6.93]	16.3 [13.1]	0.209	+1.5	4.4–7.6
Malic acid	1.32	0.023	+4.9	0.27 [0.16]	1.30 [1.60]	1.81 [1.59]	0.86 [0.67]	1.19 [1.37]	0.006	+4.5	0.7–5.3
N-Tiglylglycine	1.22	0.034	+1.7	1.55 [1.15]	2.65 [1.39]	10.9 [9.15]	10.9 [9.35]	10.8 [8.64]	0.002	+7.0	0.78–1.2
2-Hydroxyglutaric acid	1.06	0.023	+1.7	2.88 [0.99]	4.91 [3.39]	10.0 [5.90]	7.22 [3.62]	3.50 [5.73]	0.003	+3.0	0.8–52.0
Ethanol ^b	n/a	n/a	n/a	0.0	1368	5173	2852	1810	n/a	n/a	n/a

Table 1. Univariate, multivariate and descriptive statistics for the most perturbed metabolites following alcohol consumption. A total of 13 organic acid metabolites were identified as important discriminatory variables one hour after alcohol consumption ($\text{VIP} \geq 1.0$, $p \leq 0.05$ and $|\text{FC}| \geq 1.5$). Abbreviations used: [Std. Dev.], standard deviation; Cr, urinary creatinine value; HMDB, Human Metabolome Database; [WRT]₀₋₁ and [WRT]₀₋₄, Wilcoxon signed-rank test p -values showing the significance of a metabolite at times 1 and 4 hours relative to time 0; [FC]₀₋₁ and [FC]₀₋₄, fold change values for a metabolite at times 1 and 4 hours relative to time 0; n/a, not applicable for the purpose of this table. ^a13 organic acid metabolites (concentrations expressed as $\mu\text{mol}/\text{mmol Cr}$), identified through GC-MS analysis. ^bEthanol (concentrations expressed as $\mu\text{mol}/\text{mmol Cr}$), identified through an independent, targeted $^1\text{H-NMR}$ analysis of the same urine samples as used for the GC-MS analyses.

Consumed ethanol is known to be dispersed through exhaling (not measured), metabolic conversions and urinary excretion, which peaked at two hours following its consumption (Table 1). However, the results from the ASCA and unfolded PCA (Fig. 2) — which indicated that the most dramatic metabolite changes occurred one hour following the intervention — encouraged us to preferably study in detail the metabolic effects seen one hour after alcohol consumption. Of the important metabolites, hippuric acid, the phase II biotransformation product of benzoic acid^{27,35–38}, had the highest VIP value (Table 1, VIP = 19.6). The concentration of hippuric acid was already high at time 0 (948 $\mu\text{mol}/\text{mmol Cr}$), due to its being the normal excretion product of benzoic acid derived from the gut microbiome. This concentration increased to 3037 $\mu\text{mol}/\text{mmol Cr}$ one hour after the alcohol consumption, owing to the benzoic acid preservative in the flavoured water vehicle consumed with the alcohol. Hippuric acid likewise appeared to be the most important discriminating variable in the previously described effects of the vehicle-only intervention³⁵, as well as in the consecutive acute ethanol intervention study on rats, where the change in urinary hippuric acid is suggested to be due to a metabolic dysfunction of damaged liver tissue²⁷. The presence of the high concentration of hippuric acid does not influence the outcome of the analysis, since the same list of important metabolites (VIP > 1.0) is obtained after excluding hippuric acid from the data (see section 4.6 of the SI). It is also worth noting that four other gut-related urinary metabolites observed in the rat study²⁷ were also present in the urine samples from our study (see Supplementary Table S2 in section 3 of the SI), but did not appear to be discriminatory metabolites due to the alcohol intervention. Other important biotransformation products from the rat and mouse intervention studies, such as ethyl glucuronide and ethyl sulphate²⁵, were not detectable by the GC–MS methods used in the present study.

The metabolic interrelations among the remaining 12 metabolites in Table 1 correspond to several metabolic consequences of alcohol consumption. The oxidation of ethanol by alcohol dehydrogenase (ADH) creates a highly reduced cytosolic environment in hepatocytes, and favours the production of lactic acid (second-highest VIP) from pyruvic acid, resulting in downstream metabolic consequences due to pyruvic acid depletion. The reduced environment also accounts for the perturbations observed for vanillylmandelic acid (FC = −1.6), 2-hydroxybutyric (FC = +4.3) and 3-hydroxybutyric (FC = +5.9) acids, and increased urinary excretion of succinic, fumaric, malic and 2-hydroxyglutaric (derived from 2-ketoglutaric acid) acids that implicates a dysfunctional Krebs cycle. Increased excretion of N-tiglylglycine (the phase II biotransformation product of tiglyl-CoA) and 2-ethylhydracrylic acid (intermediates in the R- and S-pathways of isoleucine catabolism, respectively), as well as 3-hydroxyisobutyric acid (produced from valine catabolism) points to amino acid mobilization but inhibition of gluconeogenesis³⁹. The reason for the increased excretion of 2-hydroxyisobutyric acid (2-HIBA) remains speculative and will be discussed below.

To characterize the metabolic relationship between the 13 important variables further, we calculated Spearman’s rho correlation coefficients over the full period of the alcohol intervention, as described in section 4.4 of the Supplementary Information (SI), and shown in Fig. 3.

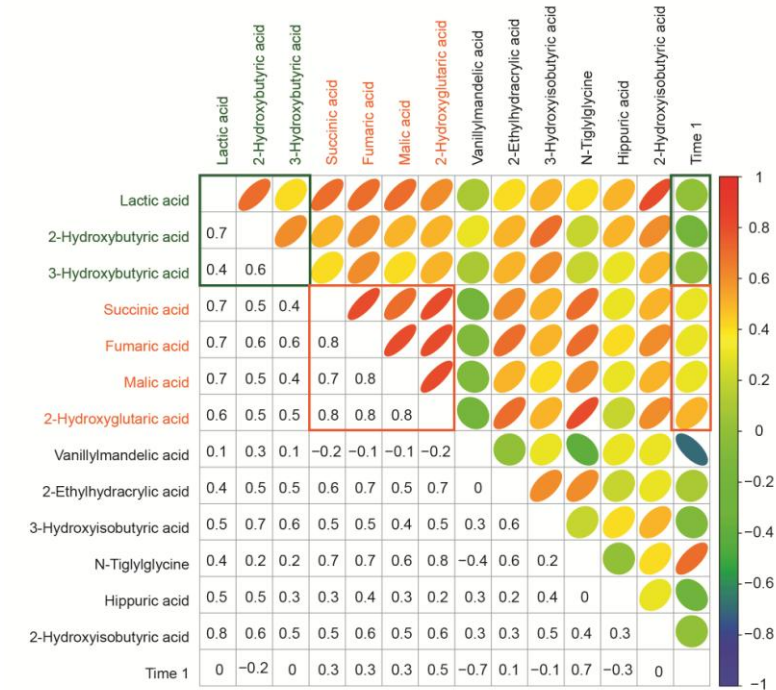


Figure 3. Correlation matrix over the full time period for the 13 metabolites listed in Table 1. Correlations extend from high positive (red; $r \geq 0.6$) through neutral (green; $-0.2 < r < 0.2$) to high negative (blue; $r \leq -0.6$). The high correlation between lactic, 2-hydroxybutyric and 3-hydroxybutyric acids and the high correlation between the Krebs cycle-associated metabolites (together with their respective time slots) are blocked in green and orange, respectively.

Important observations from the correlation analysis correspond to the metabolic interrelations: (1) the profile for vanillylmandelic acid was unique — it was the only metabolite that showed a negative correlation with time ($r = -0.7$), indicating that its urinary concentration constantly decreased over the study period, and remained very low ($1.75 \mu\text{mol}/\text{mmol Cr}$; $p = 0.034$; $\text{FC} = -8.8$) towards the end (at time 4 hours). (2) A good correlation ($r = 0.4$ to 0.7) was observed between the indicators of ketosis (lactic, 2-hydroxybutyric and 3-hydroxybutyric acids), but only in the initial phase following alcohol consumption, as indicated by their neutral correlations with time ($r = 0$ to -0.2 with time). (3) A high to very high correlation ($r = 0.6$ to 0.8) was observed between the Krebs cycle-associated metabolites, all of which also show a reasonable correlation ($r = 0.3$ to 0.5) over the full period following the intervention. (4) Likewise, N-tiglylglycine showed a high correlation with the Krebs cycle intermediates ($r = 0.6$ to 0.8), which also extended over the full study period ($r = 0.7$).

The high correlation values shown in Fig. 3 collectively substantiate a strong interrelationship between the 13 organic acid metabolites that were regarded as important due to acute alcohol consumption. The mean concentrations, and the time 4 relative to time 0 p -values ($[\text{WRT}]_{0-4}$) and fold changes ($[\text{FC}]_{0-4}$) for the 13 variables (listed in Table 1), support the conclusions about the individual and group correlations observed in Fig. 3; they also provide an alternative and additional support for the observations from the correlation study.

Inter-individual variation following acute alcohol consumption. In order to illustrate the inter-individual variation between the participants, as well as the longitudinal effect of the alcohol consumption, PCA of the data unfolded in time was performed as described in section 4.5 of the SI. The unfolding transformed the three-dimensional data (a tensor of cases, interventions and time) into a two-dimensional matrix, and thus allowed PCA to account for the longitudinally repeated measures.

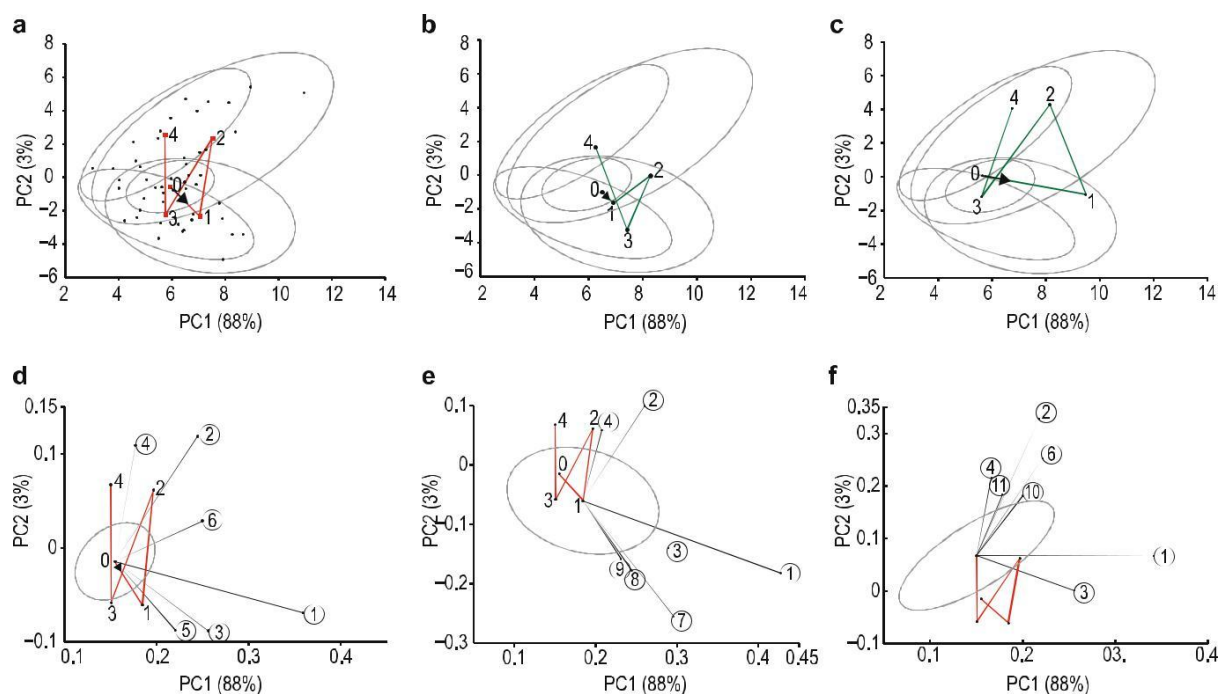


Figure 4. Unfolded PCA scores and selected bi-plots. **(a)** PC1 and PC2 of the mean of the cases studied are shown (in red), as well as the 90% confidence ellipsoids for scores of PC1 and PC2 at time 0, 1, 2, 3 and 4 hours. The centroids of the five time-dependent clusters are indicated by red squares. The direction of the trajectory linking the centroids from time 0 onwards is indicated by a black arrow. **(b)** and **(c)** The trajectories of the spectral profiles of two individual cases, illustrating the variation in the averaged and individual profiles in response to the alcohol intervention. **(d–f)** Bi-plots showing the seven most important metabolites when ranking was based on the sum of the squares of loadings of the first two components. These metabolites are more primarily responsible for the pattern in the plot, that is, causing separation in the unfolding of the PCAs, as applicable for times 0 **(d)**, 1 hour **(e)** and 4 hours **(f)**. Encircled numbers identify the metabolites: 1, hippuric acid; 2, phenylacetylglutamine; 3, 4-hydroxyphenylacetic acid; 4, indole-3-acetic acid; 5, pyrrole-2-carboxylic acid; 6, pyroglutamic acid; 7, lactic acid; 8, 2-hydroxybutyric acid; 9, fumaric acid; 10, N-cinnamoylglycine; 11, glucuronic acid.

Figure 4a–c illustrates the scores based on the first two principal components (PC1 and PC2) of the PCA model. This analysis of the unfolded data provides insight into the effect of the acute alcohol consumption over time on the 120 metabolites. The centroids of the PCA scores for each time (Fig. 4a), as well as those of two individual cases (Fig. 4b, c), illustrate the inter-individual variation in response to the intervention. The averaged and individual trajectories showed similarities in the individual responses to the alcohol consumption over time, since they all indicated a biphasic response pattern (phase 1: times 0 to 1 to 2, and phase 2: times 2 to 3 to 4). However, distinct

differences were also noted, indicated by the unique orientation and biphasic profiles of the trajectories of the individuals (Fig. 4b, c). The metabolic basis for the latter observation is illustrated, for example, by a comparison of the lactic acid excretion observed in the two cases, which progressed from 8.2 to 6.0, 74.5, 14.7 and 16.0 $\mu\text{mol}/\text{mmol}$ Cr for case 1 (Fig. 4b), but from 10.6 to 31.2, 19.8, 8.7 and 11.2 $\mu\text{mol}/\text{mmol}$ Cr for case 2 (Fig. 4c). Similar differences in the individual responses to the alcohol consumption were observed for several other metabolites as well, which cumulatively result in the unique trajectory of each individual subject — an observation that agrees with the contemporary view on genetic and metabolic individuality.

Figure 4d–f illustrates the bi-plots for the top most influential metabolites responsible for the separation in the unfolding of the PCAs. The dominant metabolites observed at time 0 were phase II biotransformation products (hippuric acid and phenylacetylglutamine) and organic acids originating from the gut microbiome (4-hydroxyphenylacetic, indole-3-acetic, pyrrole-2-carboxylic and pyroglutamic acids). This gut-derived organic acid profile dominates at all the times measured, as also indicated in Fig. 4e and 4f for times 1 and 4 hours. Notably, one hour after alcohol consumption, lactic, 2-hydroxybutyric and fumaric acids appear as additional important metabolites, in accordance with the results in Table 1 and Fig. 3. At time 3, aconitic acid is indicated as an important metabolite (figure not shown), whereas N-cinnamoylglycine and glucuronic acid (both associated with phase II biotransformation) appear to be important 4 hours after alcohol consumption. It thus appears that the endogenous detoxification mechanisms through biotransformation remained functional despite the acute alcohol consumption.

DISCUSSION

In this study, the metabolomics organic acid profile revealed significant metabolic effects of a single dose of alcohol, consumed in a well-defined vehicle by healthy, moderate-drinking males. Although their analytical constraints limit the scope of profiling studies, we concur with the view that carefully conducted studies in humans are warranted, and would provide valuable new insights into the short and long term effects of alcohol exposure, alcoholic liver disease and alcoholism, in man²⁸. The method followed here offered a convenient and sensitive approach to uncover perturbed metabolic pathways, of which we modelled the main ones as illustrated in Fig. 5. Several distinct insights deserve special attention.

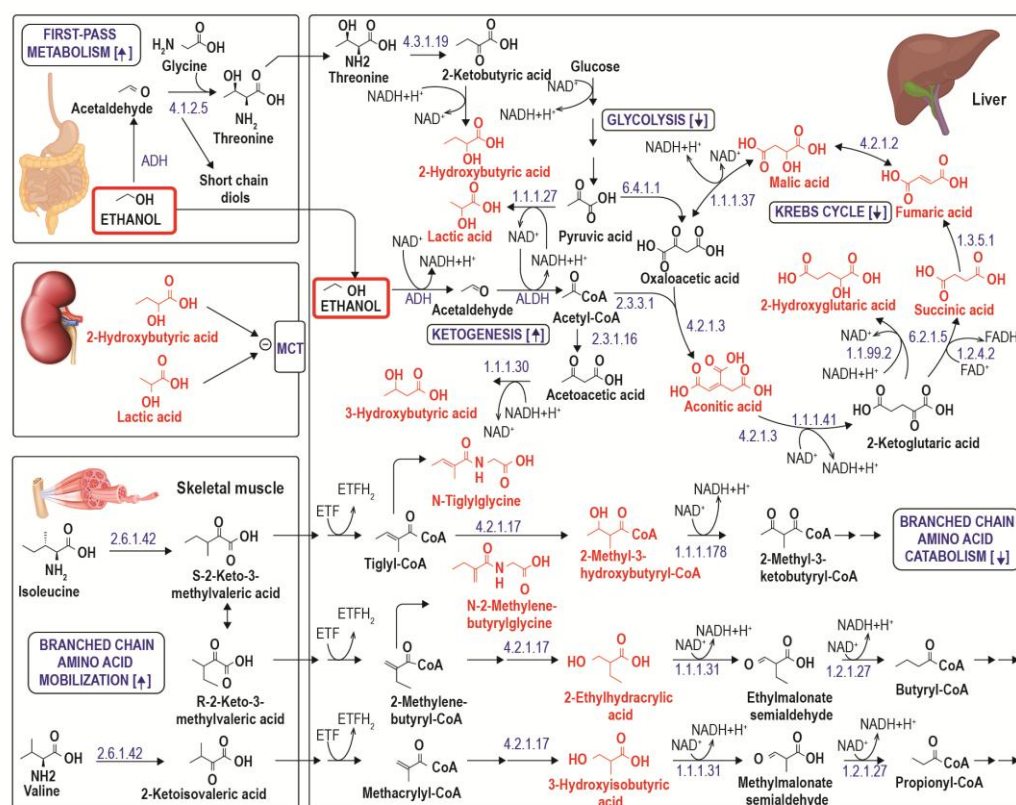


Figure 5. Proposed model indicating some important metabolic pathways affected by acute alcohol consumption. Up- or down-regulated metabolites are shown in red. Abbreviations used: MCT, monocarboxylic acid transporter; four-figure numbers (e.g. 2.6.1.42) refer to the IUBMB enzyme nomenclature (EC number); S, main catabolic pathway of isoleucine; R, minor catabolic pathway of isoleucine, which may act as a safety valve for overflow of accumulating metabolites from the S-pathway. Credit for images used: Kidney – ID 20446327 © Natis76 | Dreamstime.com; Muscle – ID 69693231 © Tigatelu | Dreamstime.com; Liver & Digestive system – ID 43552467 © Bluezace | Dreamstime.com

The first insight deals with hepatic robustness. Glycine N-acyltransferase (GLYAT)-based biotransformation was shown to play a vital background role during the intervention study. The scores and bi-plots from the unfolded PCAs (Fig. 4d–f) revealed the reasons for the observed differentiation between the groups shown in the scores plots of the unsupervised PCAs (Fig. 1a–d). The observed differentiation was caused primarily by detoxification (gut metabolites) and biotransformation (phase II conjugates) processes. Apart from the alcohol load from the intervention, the liver is simultaneously confronted with the benzoic acid load from the vehicle, which induced a significant hepatic biotransformation response³⁵. Two closely related acyl-CoA:amino acid N-acyltransferases have been characterized as operative in liver mitochondria⁴⁰. One transferase was specific for benzoyl-CoA, salicyl-CoA, and certain short straight- and branched-chain fatty acyl-CoA esters as substrates, whereas the other enzyme specifically used either phenylacetyl-CoA or indoleacetyl-CoA. It thus seems probable that the benzoyl-CoA:glycine N-acyltransferase (using glycine as the preferred acyl acceptor) was functional in the biotransformation of the vehicle-induced accumulated benzoyl-CoA, but notably also of tiglyl-CoA and 2-methylenebutyryl-CoA, which accumulated from perturbations due to the alcohol consumption. This biotransformation resulted in the formation and increased excretion of hippuric acid, N-tiglylglycine and N2-methylenebutyrylglycine (a minor metabolite increased due to the alcohol consumption — data not shown, but included in Fig. 5), respectively. The metabolomics approach thus provides insight into a background process that underscores the robustness of hepatic function in healthy individuals, notwithstanding the impact of the acute alcohol consumption — an observation that would not hold true for the liver affected by chronic alcoholism.

The second insight deals with the comprehensive metabolic perturbations caused by the acute alcohol consumption, as disclosed by the supervised PLS-DA. The metabolic reactions following alcohol consumption showed a phased response profile, to which the gut is known to contribute following alcohol consumption. In the early phase, a fraction of the alcohol consumed orally may be immediately metabolized by the gut alcohol dehydrogenase (ADH) isoforms in a process known as first-pass metabolism⁴¹. A study on acute alcohol binge drinking showed increased serum levels of bacterial products (endotoxin and bacterial 16S rDNA) derived from the gut microbiome⁴². The serum endotoxin levels rapidly increased within 30 minutes following consumption, remained elevated for 3 hours and returned to lower than baseline levels by 24 hours after alcohol intake. The early increase in 2-hydroxybutyric acid following alcohol consumption observed in the present study (Table 1, Fig. 3 and Fig. 4e — metabolite 8) is an indication of a limited contribution from first-pass metabolism in the present experimental group: threonine is anticipated to be produced from the increased intestinal acetaldehyde (the product of ethanol metabolism), and is subsequently catabolised in the liver to 2-hydroxybutyric acid, as proposed in the representation shown in Fig. 5. However, since the liver is important in alcohol metabolism, and considering its wide spectrum of lesions⁴³, we linked most of the metabolomic perturbations observed in the present study to the liver metabolism, as shown in the conceptual model in Fig. 5. Alcohol oxidation by hepatic ADH results in the reduction of NAD^+ to NADH, thereby generating a highly reduced cytosolic environment in hepatocytes. The increased NADH:NAD⁺ ratio influences the acetaldehyde–lactate dehydrogenase coupled lactic acid accumulation (ALDH–LDH = EC1.2.1.10–EC1.1.1.27). Greatly increased lactic acid excretion (Table 1, Fig. 3 and Fig. 4e — metabolite 7) is proposed as the next early hallmark of acute alcohol consumption and, together with the concomitant early increase in 3-hydroxybutyric acid (Table 1 and Fig. 3), is linked to ketogenesis. Acetoacetic acid, the precursor of 3-hydroxybutyric acid, showed a similar early increase (0.012, 0.016 and 0.044 $\mu\text{mol}/\text{mmol Cr}$ at times 0, 1 and 2 hours, respectively), and peaked at two hours after alcohol consumption (VIP = 0.004; FC = 3.64; $p = 0.02$). Taken together, first-pass metabolism and ketogenesis seem to be the immediate salvage responses to disturbed homeostasis caused by the acute alcohol consumption.

Further, an increased NADH:NAD⁺ ratio (depleted NAD⁺) has been described to influence the direction of several metabolic processes, including decreased glycolysis⁴⁴, decreased Krebs cycle^{20,45}, and decreased gluconeogenesis^{20,46}. The early increase in fumaric acid (Table 1, Fig. 3 and Fig. 4e — metabolite 9) indicates reduced Krebs cycle functionality, which persisted to the end of the 4-hour period of the present intervention study (Table 1 and Fig. 3). The notable and persistently reduced Krebs cycle functionality is further supported by the inhibition of the catabolism of all three branched-chain amino acids, indicated by the profiles of N-tiglylglycine, 2-ethylhydracrylic acid and 3-hydroxyisobutyric acid (Table 1 and Fig. 3). It is worth noting that the perturbed urinary metabolite profile had not normalized 4 hours after alcohol consumption, although the supervising physician regarded the participants sufficiently sober to permit them to leave the clinic at this time.

The third insight highlights the observations on vanillylmandelic acid. This metabolite was significantly down-regulated by depleted NAD⁺, and remained so over the full 4-hour period (Table 1 and Fig. 3). The down-regulation of vanillylmandelic acid reflects the upstream inhibition in the catabolism of the catecholamines noradrenaline and adrenaline, which occurs due to the inhibition of the NAD⁺-dependent aldehyde dehydrogenase-catalysed reactions (EC1.2.1.3 and EC1.2.1.5, respectively). The known relative rise in adrenaline and the delayed increase in noradrenaline concentrations after alcohol consumption is a physiological effect of the inhibition of these enzymes⁴⁷, which, together with the corticotrophin-releasing factor, may contribute to negative affective states and relapse vulnerability during alcohol abstinence⁴⁸. The observations on vanillylmandelic acid may also be important for other alcohol related studies. In a study on alcoholic patients, higher levels of urinary homovanillic acid (an up-stream metabolite of the catecholamine pathway) were observed in patients with a A1 allele in their genotype, when compared to patients homozygous for the TaqIA2 allele⁴⁹. We observed that homovanillic acid peaked two hours after acute alcohol consumption (VIP = 0.129; FC = 1.13; $p = 0.44$). These findings suggest that determination of urinary levels of catecholamine end-products should be included in alcohol-related studies, since variance in their levels could relate to a polymorphism, and could affect addictive behaviour.

As a last insight, 2-hydroxyisobutyric acid (2-HIBA) was indicated in this study to be an important and significant organic acid indicator regarding the metabolic consequences of acute alcohol consumption.

2-Hydroxybutyric acid and 2-HIBA are normally reabsorbed in the kidney by a monocarboxylic acid transporter (MCT). This transporter is, however, also responsible for transporting lactic acid between the urine and the blood. After alcohol consumption, the increased lactic acid contributes to the saturation of these transporters. Consequently, not all of the 2-hydroxybutyric acid and 2-HIBA is reabsorbed, leading to their increased excretion in the urine. However, the conventional view is that urinary 2-HIBA is a non-metabolite, obtained from environmental exposure to tertiary-butylacetate (a commercial solvent used in industrial coatings and cleaners) and certain gasoline additives^{50–52}. But things change. New research casts a valuable light on 2-HIBA, suggesting that it appears to be an important metabolite with potential value for health and disease — 2-HIBA is reported to be up-regulated in known human disorders such as chronic kidney disease⁵³, diabetes mellitus^{54,55}, human gastric cancer⁵⁶, fibromyalgia⁵⁷, myocardial infarction^{58–60}, and a number of inborn errors of metabolism^{61,62}. In the present study, alcohol induced significant up-regulation of 2-HIBA in the first two hours after consumption (Table 1: VIP = 1.39; FC = +1.8; p = 0.034). The very high correlation of 2-HIBA with lactic acid (r = 0.8), and high correlation with 2-hydroxybutyric and fumaric acids (r = 0.6), strongly supports the view that 2-HIBA is an endogenous metabolite. However, considering the diversity of the conditions associated with the up-regulation of 2-HIBA, a common metabolic origin remains elusive. Most notably, MS analyses of peptide fragments from human histones recently identified a new type of histone mark, namely lysine 2-hydroxyisobutyrylation, which is conserved, widely distributed, has a high stoichiometry, and induces large structural change in histones⁶³. Histone acetylation was first identified more than 50 years ago⁶⁴, and has since become known as wide-ranging posttranslational modifications, linked to a variety of processes, including transcription, DNA replication, and DNA damage. A comprehensive catalogue of histone modifications and their proposed functional consequences⁶⁵ includes 2-hydroxyisobutyrylation of lysine side chains, amongst others, for which a rather limited function has been determined. Histone modifications are proposed to present a large opportunity in the years to come to gain insights into chromatin biology, epigenetic events and their biological consequences, which should include the proposed alcoholism-related involvement of 2-HIBA⁶⁶ in the modification of histones^{67,68}.

Metabolomics studies, however, have limitations: (1) GC–MS analysis, our method of choice, produces data that are noisier than other metabolomics data types, such as NMR or triple quadrupole liquid chromatography–mass spectrometry (LC–QQQ–MS). This may be overcome by using more experimental subjects (but their number was limited in this study due to ethical considerations), or by using replicates of all samples (rather than only random repeat samples as used in our measurement design). (2) Although it would be useful, quantifying metabolomics data absolutely is often difficult to execute accurately, is relatively expensive, and is rarely done in diagnostic or intervention study designs as presented here. As an acceptable alternative, therefore, we expressed our metabolite concentrations as values relative to an internal standard. (3) Concomitant plasma or serum analyses did not form part of this experimental design. Therefore, the metabolomic variables considered in this investigation were generated from only organic acid analyses of urine samples, and were modelled on theoretical grounds on metabolic pathways of the gut, muscle and liver (Fig. 5). Given the intoxicating influence of alcohol, metabolite levels in the brain are also important, but could also not be considered here due to the biofluid used. However, as baseline, the perturbations of urinary organic acid profiles provide a prime and feasible way for clinical tests to screen for individual response types, as well as for other relevant aspects of acute alcohol exposure.

Conclusions. Taken together, our findings indicate that metabolomics provided a systematic and standardized method for detecting a range of metabolic responses over time, not previously described comprehensively for acute alcohol consumption. These findings open avenues for potentially important future investigations in alcohol research: (1) genotype-based selection of individuals in follow-up alcohol intervention studies is advised, given the clear inter-individual responses to alcohol consumption; (2) low values of urinary vanillylmandelic acid may be an indicator of a binge drinking or acute alcohol consumption episode in seemingly non-intoxicated individuals; and (3) the striking presence of 2-HIBA supports the emerging new paradigm of 2-HIBA being an important endogenous metabolite. Moreover, detailed studies on the biological origin of 2-HIBA, as well as on its perceived gene-modification role through lysine 2-hydroxyisobutyrylation of histones may take us one step closer to understanding the personalized responses to acute alcohol consumption and the perceived epigenetic changes that are induced. All in all, we concur that acute alcohol consumption studies broaden insights on significant adverse health effects of alcohol even in healthy individuals⁴². These insights will help researchers to define novel approaches to treat or ameliorate alcohol-induced disability, organ damage and morbidity.

MATERIALS AND METHODS

Intervention study design. The experimental group consisted of 12 clinically selected healthy males (aged 20–24 years), who admitted to consuming alcohol at a moderate, social level (baseline alcohol consumption was defined by the participants' declared levels of drinking). None of the participants used any medication. They were asked to refrain from vitamins, minerals, and other supplementation, and to follow a similar dietary and lifestyle pattern for the duration of the study. The experimental interventions were performed at the Health Clinic of North-West University under controlled conditions. A medical doctor and nurse were present during the period of intervention, and all participants could leave the premises only after approval by the medical doctor. The protocol was approved by the Health Sciences Ethical Committee of North-West University (ethical approval number: NWU-00045-12-S1), conducted in accordance with guidelines for good clinical practice, and all participants provided informed written consent to the research protocol (an example of the informed consent form is included in section 5 of the SI). The protocol was registered as a clinical trial on 3 November 2017 under the Pan African Clinical Trial Registry (registration number: PACTR201711002748255), under the title: *A metabolomics investigation on experimental interventions of acute alcohol consumption*.

The experiments were conducted on two Saturday mornings between 08:00 and 12:00. All participants were required to abstain from alcohol consumption for at least 48 hours preceding the experiment, and to abstain from breakfast on the days of the experiments (that is, to remain in an overnight fasted state). On the first Saturday, half of the participants (randomly selected) were given 500 mL lemon-flavoured sparkling water as vehicle only (contents: fructose and citric acid flavouring; sodium benzoate preservative; sodium cyclamate, aspartame, acesulfame K sweeteners; vitamin C). The other half of the participants received the same quantity of the vehicle, as well as a predefined quantity of alcohol — 1.5 mL alcohol per kilogram body mass⁶⁹. The alcohol used was triple-distilled vodka: 43% alcohol. On the second Saturday, the participants received the alternate intervention to the one received on the first Saturday. On both days the participants were also provided with 1.5 L bottled water, which was the only substance that could be consumed over the 4-hour period of the experiment. Differences in the concentrations of urinary metabolites due to variation in water consumption between the participants were accounted for by determining the creatinine concentration of each sample, and expressing the concentrations of all the quantified metabolites as μmol metabolite/mmol creatinine. Initial (time 0) urine samples were collected just prior to the intervention. Subsequent urine samples were collected at 1, 2, 3 and 4 hours after the start of the experiment. This gave a total of five urine samples from each participant for each intervention. Time 0 urine samples served as controls for each participant, with longitudinal data being compared accordingly.

Sample handling. After collection, each urine sample was divided into aliquots and stored at -80°C . Once all the urine samples had been collected, a 1 mL aliquot of each sample was thawed and combined to prepare a pooled quality control (QC) sample. This QC sample was divided into aliquots and once again stored at -80°C . Another 1 mL aliquot of each urine sample was used for creatinine determination, performed by an external pathology service.

Metabolomics workflow. The workflow of the intervention study (Fig. 6) started with the generation of time-dependent quantitative metabolomics data, progressed to the application of various models of statistical analysis, which eventually led to the biological interpretation of the effect of the interventions on the group as well as on individual cases. The samples collected prior to alcohol (or vehicle) consumption were used as controls for the subsequent hourly samples collected after the interventions. The effect of the vehicle-only intervention was previously described³⁵, and will not be discussed here.

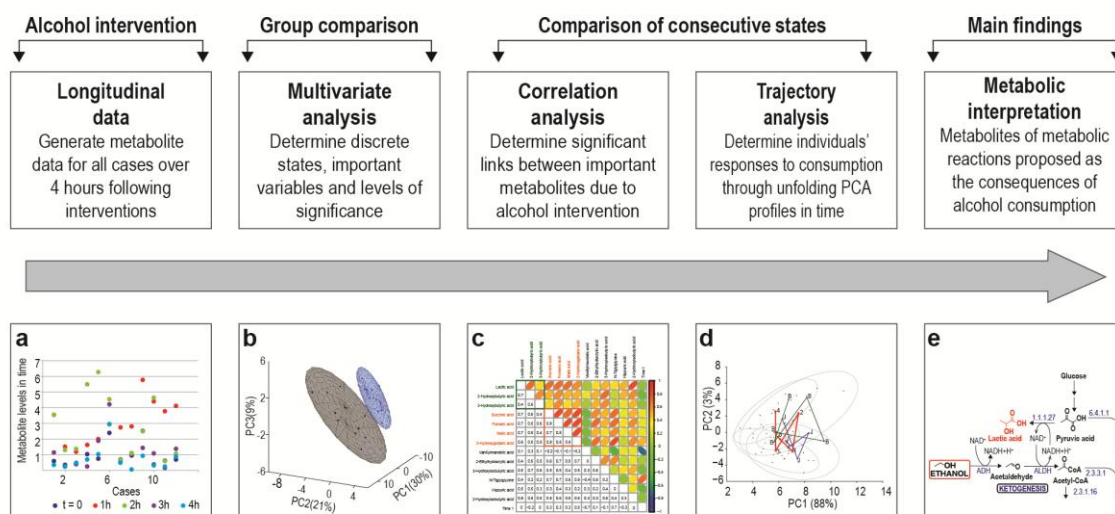


Figure 6. Representation of the metabolomics workflow to investigate the effect of acute alcohol consumption. **(a)** Longitudinal data were generated through the GC–MS analyses for 120 metabolites across five hourly intervals. **(b)** Two metabolic states — before alcohol consumption (time 0) and after alcohol consumption (times 1 to 4 hours) — were compared through multivariate, as well as univariate, analyses to establish the levels of significance of the observed differences. **(c)** Correlation analysis was used to indicate the relationship between important metabolites. **(d)** Unfolded PCA was applied to the data that emerged in time to summarize the subsequent variability, and, more importantly, between individuals, due to the alcohol effect, as well as for the identification of the metabolites responsible for this variability. **(e)** Construction of a global metabolite profile from the combined results provided the framework for the discussion on the effects of acute alcohol consumption on the subjects' metabolism.

Measurement design. For the generation of data through GC–MS analysis, each participant's samples were analysed in a separate batch. Each batch included the participant's 10 urine samples (S), repeat samples (R) and quality control samples (Q), and each batch was constructed and analysed as follows:

QQQQQSSVSVRAQS SVRAQSASASARVQSASARVQQ
where:

Q	= Quality control sample	[Total = 11]
S _V	= Randomly selected vehicle sample	[Total = 5]
S _A	= Randomly selected alcohol sample	[Total = 5]
R _A	= Randomly selected alcohol repeat sample	[Total = 2]
R _V	= Randomly selected vehicle repeat sample	[Total = 2].

QC samples were included to estimate any batch effect or other interfering analytical aspects. The data from the five QC samples at the start of each batch were used to condition the column and were excluded from further analysis. Repeat samples were included to determine their repeatability across the batch — the first two repeat samples in each batch were repeats of experimental samples from later in the batch, and the last two repeat samples were repeats of experimental samples from earlier in the batch. The total GC–MS running time for each batch was approximately 21 hours.

Organic acid extraction and GC–MS analysis. A 5 mL aliquot of each urine sample was used for organic acid extraction before GC–MS analysis, as described previously⁷⁰. All experimental, QC and repeat samples were prepared and derivatized individually and in the same way. Full details of the method are presented in section 1 of the SI. An excerpt of the data used for the alcohol intervention is shown in Supplementary Table S1 (section 2 of the SI). Metabolite concentrations are expressed as $\mu\text{mol}/\text{mmol}$ creatinine relative to an internal standard (4-phenylbutyric acid).

Quantitative urinary alcohol excretion. Urinary ethanol is not detectable by the GC–MS method used here due to its high volatility, but was quantified in all samples obtained from the vehicle and alcohol interventions by means of NMR spectroscopy. The data indicated that ethanol excretion peaked at 2 hours following alcohol consumption (data not shown).

Variable identification, classification and reduction. The QC samples were used to identify and classify a list of representative variables in all the urine samples. Following untargeted GC–MS data generation, a total of 172 variables were detected through AMDIS, excluding the 2 internal standards used. Several of the variables were present in concentrations just above the detection limit, and were not observed in all QC samples. The Human Metabolome Database (www.hmdb.ca) was used as the reference for the biological description of each feature, and as the basis for classifying the variables. Information on variables not included in the HMDB (e.g. exogenous substances or artifacts formed during the derivatization reactions) was obtained from other established chemical databases or from the literature; failure with this resulted in the classification of 7 variables as “no annotation”. Details of the 172 variables are summarized in Supplementary Table S2, and exclusion criteria for the determination of metabolite relevance are described in section 3 of the SI. From this protocol, 120 metabolites were identified and used for the statistical evaluation of the effect of alcohol consumption.

Statistical analysis

Identification of important organic acid metabolites. Metabolites causing the separations were regarded as important if they varied substantially between the samples collected prior to alcohol consumption and those collected one hour after alcohol consumption. Changes in metabolite levels were ranked based on their multivariate VIP values (Variable Importance in Projection), fold change (FC) values and non-parametric Wilcoxon signed-rank test (WRT) p -values. The selection criteria were: $\text{VIP} \geq 1.0$, $\text{WRT } p \leq 0.05$ and $|\text{FC}| \geq 1.5$. The aim with the selection was for a deeper understanding of the dominant biological changes rather than to model the observed data.

Multivariate statistical analysis. Two modes of multivariate statistical analyses were applied to the metabolomics data generated during this intervention experimental design: (1) cross-sectional analysis of time points using traditional multivariate methods to compare two groups (PCA and PLS-DA); and (2) longitudinal analysis performed across all times (ASCA^{34,71} and unfolded PCA⁷²). Details of these methods were previously described³⁵.

The statistical analyses indicated that we should achieve acceptable power (0.8) for a large effect size (0.9) given 12 paired observations and a 5% significance level. We should therefore be able to identify large differences between two factor levels (that is, between two points in time or two interventions) for a single variable (see SI section 4.1 for details). Prior to statistical analysis, the data were pre-processed by: (i) treatment of zero-valued observation; and (ii) transformation and scaling. This is explained in greater detail in section 4.2 of the SI. PCA was used to project the observed data to new spaces that maximizes the variation along fewer hyperplanes while not taking the group membership into consideration. PLS-DA was applied to build models to predict group membership, by projecting the variance in the observed data measured and the membership to new spaces. PLS-DA was used as a supervised method to rank and select the metabolites most changed by the intervention. The significance of these changes was established through univariate analysis using the Wilcoxon signed-rank test and fold change ratios. The figures shown in Fig. 1 do not take into account the paired nature of the data. However, multi-level PCA and PLS-DA were also performed, which cater specifically for repeated measures. The multi-level results are closely related to those reported here, and are included in sections 4.3 and 4.5 of the SI.

Data availability. The full data set is given in Excel format as part of the SI (available online).

References

- World Health Organization (2014) Global status report on alcohol and health 2014. Available at: www.who.int/substance_abuse/publications/global_alcohol_report/en/. Accessed July 16, 2017.
- Shield, K. D., Parry, C. & Rehm, J. Chronic diseases and conditions related to alcohol use. *Alcohol Res.* **35**(2), 155–173 (2013).
- Blakemore, S.-J. & Choudhury, S. Development of the adolescent brain: implications for executive function and social cognition. *J. Child Psychol. Psychiatry* **47**, 296–312 (2006).
- Lenroot, R. K. & Giedd, J. N. Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. *Neurosci. Biobehav. Rev.* **30**, 718–729 (2006).
- White, A., Matthews, D. & Best, P. Ethanol, memory, and hippocampal function: a review of recent findings. *Hippocampus* **10**, 88–93 (2000).
- Fein, G. *et al.* Cortical and subcortical volumes in adolescents with alcohol dependence but without substance or psychiatric comorbidities. *Psychiatry Res.: Neuroimaging* **214**(1), 1–8 (2013).
- Guerrini, I., Quadri, G. & Thomson, A. D. Genetic and environmental interplay in risky drinking in adolescents: a literature review. *Alcohol Alcohol.* **49**, 138–142 (2014).
- Mäkelä, K. & Mustonen, H. Relationships of drinking behaviour, gender and age with reported negative and positive experiences related to drinking. *Addiction* **95**, 727–736 (2000).
- O’Neil, S. *et al.* Brief intervention to prevent hazardous drinking in young people aged 14–15 in a high school setting (SIPS JR-HIGH): study protocol for a randomized controlled trial. *Trials* **13**(1), 166 (2012).
- Spanagel, R. *et al.* An integrated genome research network for studying the genetics of alcohol addiction. *Addict. Biol.* **15**(4), 369–379 (2010).
- Rietschel, M. & Treutlein, J. The genetics of alcohol dependence. *Ann. N. Y. Acad. Sci.* **1282**(1), 39–70 (2013).
- Morozova, T. V., Mackay, T. F. & Anholt, R. R. Genetics and genomics of alcohol sensitivity. *Mol. Genet. Genomics* **289**(3), 253–269 (2014).
- Nahon, P. *et al.* Identification of serum proton NMR metabolomic fingerprints associated with hepatocellular carcinoma in patients with alcoholic cirrhosis. *Clin. Cancer Res.* **18**(24), 6714–6722 (2012).
- Rachakonda, V. *et al.* Serum metabolomic profiling in acute alcoholic hepatitis identifies multiple dysregulated pathways. *Plos one* **9**(12), e113860 (2014).
- Amathieu, R. *et al.* Serum 1H-NMR metabolomic fingerprints of acute-on-chronic liver failure in intensive care unit patients with alcoholic cirrhosis. *Plos one* **9**(2), e89230 (2014).
- Manna, S. K. *et al.* UPLC–MS-based urine metabolomics reveals indole-3-lactic acid and phenyllactic acid as conserved biomarkers for alcohol-induced liver disease in the Ppara-null mouse model. *J. Proteome Res.* **10**, 4120–4133 (2011).
- Gill, J. S. Reported levels of alcohol consumption and binge drinking within the UK undergraduate student population over the last 25 years. *Alcohol Alcohol.* **37**(2), 109–120 (2002).
- Mathurin, P. & Deltenre, P. Effect of binge drinking on the liver: an alarming public health issue? *Gut* **58**, 613–617 (2009).
- Askgaard, G., Grønbaek, M., Kjær, M. S., Tjønneland, A. & Tolstrup, J. S. Alcohol drinking pattern and risk of alcoholic liver cirrhosis: a prospective cohort study. *J. Hepatol.* **62**(5), 1061–1067 (2015).
- Zakhari, S. & Li, T.-K. Determinants of alcohol use and abuse: impact of quantity and frequency patterns on liver disease. *Hepatology* **46**, 2032–2039 (2007).
- Szabo, G. Gut–liver axis in alcoholic liver disease. *Gastroenterology* **148**(1), 30–36 (2015).
- Petit, G., Muraige, P., Kornreich, C., Verbanck, P. & Campanella, S. Binge drinking in adolescents: a review of neurophysiological and neuroimaging research. *Alcohol Alcohol.* **49**(2), 198–206 (2013).
- Rehm, J., Taylor, B. & Room, R. Global burden of disease from alcohol, illicit drugs and tobacco. *Drug Alcohol Rev.* **25**, 503–513 (2006).
- Taylor, B. *et al.* The more you drink, the harder you fall: a systematic review and meta-analysis of how acute alcohol consumption and injury or collision risk increase together. *Drug Alcohol Depend.* **110**(1), 108–116 (2010).
- Gika, H. G. *et al.* Investigation of chronic alcohol consumption in rodents via UHPLC–MS based metabolite profiling. *J. Chromatogr. A* **1259**, 128–137 (2012).
- Szabo, G., Bala, S., Petrasek, J. & Gattu, A. Gut–liver axis and sensing microbes. *Dig. Dis.* **28**(6), 737–744 (2010).
- Gao, X. *et al.* GC/MS-based urinary metabolomics reveals systematic differences in metabolism and ethanol response between Sprague–Dawley and Wistar rats. *Metabolomics* **7**(3), 363–374 (2011).
- Gika, H. G. & Wilson, I. D. Global metabolic profiling for the study of alcohol-related disorders. *Bioanalysis* **6**(1), 59–77 (2014).
- Gika, H. G. & Wilson, I. D. Metabolic profiling approaches for biomarkers of ethanol intake in *Molecular Aspects of Alcohol and Nutrition* (ed Patel, V. B.) 213–222 (Chapter 17) (Elsevier, 2016).
- Plant, M., Plant, M. & Vernon, B. Ethics, funding and alcohol research. *Alcohol Alcohol.* **31**(1), 17–25 (1996).
- Gibney, M. J. *et al.* Metabolomics in human nutrition: opportunities and challenges. *Am. J. Clin. Nutr.* **82**, 497–503 (2005).
- Wishart, D. S. Metabolomics: applications to food science and nutrition research. *Trends Food Sci. Tech.* **19**, 482–493 (2008).
- Antti, H. *et al.* Batch statistical processing of 1NMR-derived urinary spectral data. *J. Chemometrics* **16**, 461–468 (2002).
- Smilde, A. K. *et al.* ANOVA simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics* **21**, 3043–3048 (2005).
- Irwin, C. *et al.* Contribution towards a metabolite profile of the detoxification of benzoic acid through glycine conjugation: an intervention study. *Plos one* **11**(12), e0167309 (2016).
- Roberts, M. ‘A big night out’: young people’s drinking, social practice and spatial experience in the ‘liminoid’ zones of English night-time cities. *Urban Stud.* **52**(3), 571–588 (2015).
- Conti, A. & Bickel, M. H. History of drug metabolism: discoveries of the major pathways in the 19th century. *Drug Metab. Rev.* **6**(1), 1–50 (1977).
- Badenhorst, C. P. S., Erasmus, E., Van der Sluis, R., Nortje, C. & Van Dijk, A. A. A new perspective on the importance of glycine conjugation in the metabolism of aromatic acids. *Drug Metab. Rev.* **46**(3), 343–361 (2014).
- Costantini, P., Chernyak, B. V., Petronilli, V. & Bernardi, P. Modulation of the mitochondrial permeability transition pore by pyridine nucleotides and dithiol oxidation at two separate sites. *J. Biol. Chem.* **271**(12), 6746–6751 (1996).
- Nandi, D. L., Lucas, S. V. & Webster, L. T. Jr. Benzoyl-coenzyme A:glycine N-acyl-transferase and phenylacetyl-coenzyme A:glycine N-acyltransferase from bovine liver mitochondria. *J. Biol. Chem.* **254**, 7230–7237 (1979).
- DiPadova, C., Worner, T. M., Julkunen, R. J. & Lieber, C. S. Effects of fasting and chronic alcohol consumption on the first-pass metabolism of ethanol. *Gastroenterology* **92**(5), 1169–1173 (1987).
- Bala, S., Marcos, M., Gattu, A., Catalano, D. & Szabo, G. Acute binge drinking increases serum endotoxin and bacterial DNA levels in healthy individuals. *Plos one* **9**(5), e96864 (2014).
- Louvet, A. & Mathurin, P. Alcoholic liver disease: mechanisms of injury and targeted treatment. *Nat. Rev. Gastroenterol. Hepatol.* **12**(4), 231–242 (2015).
- Bradford, B. U. & Rusyn, I. Swift increase in alcohol metabolism (SIAM): understanding the phenomenon of hypermetabolism in liver. *Alcohol* **35**, (13–17) (2005).
- Hawkins, R. D. & Kalant, H. The metabolism of ethanol and its metabolic effects. *Pharmacol. Rev.* **24**(1), 67–157 (1972).
- Lieber, C. S. Ethanol metabolism, cirrhosis and alcoholism. *Clin. Chim. Acta* **257**, 59–84 (1997).
- Ireland, M. A., Van Dongen, R., Davidson, L., Beilin, L. J. & Rouse, I. L. Acute effects of moderate alcohol consumption on blood pressure and plasma catecholamines. *Clin. Sci.* **66**, 643–648 (1984).

48. Becker, H. C. Effects of alcohol dependence and withdrawal on stress responsiveness and alcohol consumption. *Alcohol Res.* **34**(4), 448–458 (2012).
- Ponce, G. *et al.* IDRD2 TaqIA polymorphism is associated with urinary homovanillic acid levels in a sample of Spanish male alcoholic patients. *Neurotox. Res.* **6**(5), 373–377 (2004).
50. Bus, J. S. *et al.* Human health screening level risk assessments of tertiary-butyl acetate (TBAC): calculated acute and chronic reference concentration (RfC) and hazard quotient (HQ) values based on toxicity and exposure scenario evaluations. *Crit. Rev. Toxicol.* **45**(2), 142–171 (2015).
51. Dekant, W., Bernauer, U., Rosner, E. & Amberg, A. Toxicokinetics of ethers used as fuel oxygenates. *Toxicol. Lett.* **124**, 37–45 (2001).
52. Amberg, A., Rosner, E. & Dekant, W. Toxicokinetics of methyl tert-butyl ether and its metabolites in humans after oral exposure. *Toxicol. Sci.* **61**, 62–67 (2001).
53. Mutsaers, H. A. M. *et al.* Optimized metabolomic approach to identify uremic solutes in plasma of stage 3–4 chronic kidney disease patients. *Plos one* **8**(8), e71199 (2013).
54. Zhang, A.-H., Qiu, S., Xu, H.-Y., Sun, H. & Wang, X.-J. Metabolomics in diabetes. *Clin. Chim. Acta* **429**, 106–110 (2014).
55. Li, X. *et al.* Comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry for metabolomics: biomarker discovery for diabetes mellitus. *Anal. Chim. Acta* **633**, 257–262 (2009).
56. Wang, H. *et al.* Tissue metabolic profiling of human gastric cancer assessed by ¹H NMR. *BMC Cancer* **16**, 371 (2016).
57. Malatji, B. G. *et al.* A diagnostic biomarker profile for fibromyalgia syndrome based on an NMR metabolomics study of selected patients and controls. *BMC Neurol.* **17**, 88–102 (2017).
58. Lee, J. *et al.* LC/MS-based polar metabolite profiling reveals gender differences in serum from patients with myocardial infarction. *J. Pharm. Biomed. Anal.* **115**, 475–486 (2015).
59. Surekha, R. H. *et al.* Oxidative stress and total anti oxidant status in myocardial infarction. *Singapore Med. J.* **48**(2), 137–142 (2007).
60. Ali, M. A. & Konishi, T. Enhancement of hydroxyl radical generation in the Fenton reaction by alpha-hydroxy acid. *Biochem. Mol. Biol. Int.* **46**(1), 137–145 (1998).
61. Goodman, S. I., Mccabe, E. R., Fennessey, P. V. & Mace, J. W. Multiple acyl-CoA dehydrogenase deficiency (glutaric aciduria type II) with transient hypersarcosinemia and sarcosinuria; possible inherited deficiency of an electron transfer flavoprotein. *Pediatr. Res.* **14**(1), 12–17 (1980).
62. Tanaka, K., Hine, D. G., West-Dull, A. & Lynn, T. B. Gas-chromatographic method of analysis for urinary organic acids. I. Retention indices of 155 metabolically important compounds. *Clin. Chem.* **26**(13), 1839–1846 (1980).
63. Dai, L. *et al.* Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nat. Chem. Biol.* **10**(5), 365–370 (2014).
64. Allfrey, V. G., Faulkner, R. & Mirsky, A. E. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci.* **51**, 786–794 (1964).
65. Zhao, Y. & Garcia, B. A. Comprehensive catalog of currently documented histone modifications. *Cold Spring Harb. Perspect. Biol.* **7**(9), a025064 (2015).
66. You, C., Zhang, H., Sakharkar, A. J., Teppen, T. & Pandey, S. C. Reversal of deficits in dendritic spines, BDNF and Arc expression in the amygdala during alcohol dependence by HDAC inhibitor treatment. *Int. J. Neuropsychopharmacol.* **17**, 313–322 (2014).
67. Volmar, C. H. & Wahlestedt, C. Histone deacetylases (HDACs) and brain function. *Neuroepigenetics* **1**, 20–27 (2015).
68. Li, Z., Bridges, B., Olson, J. & Weinman, S. A. The interaction between acetylation and serine-574 phosphorylation regulates the apoptotic function of FOXO3. *Oncogene* **36**(13), 1887–1898 (2017).
69. Numminen, H., Syrjälä, M., Benthin, G., Kaste, M. & Hillborn, M. The effect of acute ingestion of a large dose of alcohol on the hemostatic system and its circadian variation. *Stroke* **31**(6), 1269–1273 (2000).
70. Mason, S. *et al.* A putative urinary biosignature for diagnosis and follow-up of tuberculous meningitis in children: outcome of a metabolomics study disclosing host–pathogen responses. *Metabolomics* **12**, 1–16 (2016).
71. Zwanenburg, G., Hoefsloot, H. C. J., Westerhuis, J. A., Jansen, J. J. & Smilde, A. K. ANOVA-principal component analysis and ANOVA-simultaneous component analysis: a comparison. *J. Chemom.* **25**, 561–567 (2011).
72. Villez, K., Steppe, K. & De Pauw, D. J. W. Use of unfold PCA for on-line plant stress monitoring and sensor failure detection. *Biosystems Engineering* **103**, 23–24 (2009).

Acknowledgements

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 TIA. The funder had no role in designing the study, data collection and analysis, decision to publish, or preparation, writing or editing of the manuscript.

Author Contributions

Conceptualization: C.J.R., J.A.W. Data curation: C.I., S.M. M.v.R., J.A.W. Formal analysis: C.I. Funding acquisition: C.J.R. Investigation: C.I., S.M. Methodology: C.J.R., J.A.W. Project administration: C.J.R., J.A.W. Software: J.A.W., M.v.R. Supervision: C.J.R., J.A.W. Writing – original draft: C.I., M.v.R., S.M., L.J.M., J.A.W., C.J.R., R.A.W. Writing – review & editing: C.I., M.v.R., S.M., L.J.M., J.A.W., C.J.R., R.A.W.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-24128-1>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

GC–MS-based urinary organic acid profiling reveals multiple dysregulated metabolic pathways following experimental acute alcohol consumption — Supplementary Information

Cindy Irwin, Lodewyk J. Mienie, Ron A. Wevers, Shayne Mason, Johan A. Westerhuis, Mari van Reenen, Carolus J. Reinecke

CONTENTS:

1. Sample preparation, organic acid extraction and GC–MS analysis
2. Data used for the alcohol intervention study: Supplementary Table S1 — An excerpt of the data used for the alcohol intervention. The full data set is given as a separate file in Excel format.
3. Details of the 172 identified features: Supplementary Table S2 — Separation and classification details of the 172 identified features
4. Statistical Methods & Additional Results
 - 4.1 Power
 - 4.2 Data pre-processing
 - 4.3 Statistical methods for variable selection
 - 4.4 Correlation analysis
 - 4.5 Statistical methods for repeated measures data
 - 4.6 The vehicle/hippuric acid effect
 - 4.7 Statistical software
5. Example of informed consent form
6. References

1. Sample preparation, organic acid extraction and GC–MS analysis

1.1 Extraction and derivatization of organic acids

The 5 mL aliquots used for the analyses were thawed at room temperature and vortexed before use.

The volume of urine used for the organic acid analysis was based on the creatinine value of each of the samples:

- For creatinine values higher than 100 mg% use 0.5 ml urine
- For creatinine values less than 100 mg% but higher than 5 mg% use 0.5 ml urine
- For creatinine values less than 5 mg% but higher than 2 mg% use 2 ml urine
- For creatinine values less than 2 mg% use 3 ml urine.

The correct volume of urine was transferred to a 6 ml Kimax tube and 6 drops of 5M HCl was added to adjust the pH of the urine to below 2.

The volume of internal standard (IS) added was determined by multiplying the creatinine value (in mg%) with five times the volume of urine used. A mix of malonic acid (RT ~14 minutes) and 4-phenylbutyric acid (RT ~20 minutes) was used as internal standard. The IS was prepared by dissolving 52.5 mg of each in 50 ml milli-Q water. The solution was sonicated for 30 minutes to ensure that all compounds were properly dissolved. 4-Phenylbutyric acid was used owing to its absence in normal urine and in known pathological conditions. In addition, it elutes almost in the middle of the organic acid profile and theoretically co-elutes with very few, if any, other organic acids. Malonic acid was added as a second IS reference peak in the GC chromatogram.

The first extraction was done by adding 6 ml of ethylacetate to each of the samples. The tubes were capped and checked for leakage by inverting the tubes before being mixed on a rotary wheel for 30 minutes. The samples were then centrifuged at 3000 rpm for 3 minutes and the organic phase (top phase) was aspirated into a clean 6 ml Kimax tube.

The second extraction was done by adding 3 ml of diethylether to the aqueous phase (lower phase) and the tubes were once again capped and checked for leakage. The samples were then mixed on the rotary wheel for 10 minutes and centrifuged at 3000 rpm for 3 minutes. The organic phase was aspirated and added to the ethylacetate phase and the aqueous phase was discarded.

Approximately 2 ml of anhydrous sodium sulphate (Na_2SO_4) was added to each sample and the tubes were capped and vortexed to mix. Proper dispersion of the Na_2SO_4 ensures that all the water is removed from the organic phase. The samples were then centrifuged at 3000 rpm for 1 minute and the organic phase was decanted into a clean 3 ml Kimax tube. The samples were finally evaporated to dryness in a heating block at 37°C under nitrogen gas for approximately 45 minutes.

Using a Hamilton syringe, the derivatization reagents O-bis(trimethylsilyl)trifluoroacetamide (BSTFA), trimethylchlorosilane (TMCS) and pyridine were added to the dried samples. The volume of BSTFA added was determined by multiplying the creatinine value (in mg%) with three times the volume of urine used. The volume of TMCS and pyridine added was determined by multiplying the creatinine value (in mg%) with 0.6 times the volume of urine used. The tubes were capped and incubated at 70°C for 45 minutes.

The derivatized samples were then transferred to 1.5 ml GC–MS vials with inserts, capped, and placed in the GC–MS autosampler for GC–MS analysis. Blank (hexane), QC and repeat samples were included in the injection sequence, as described in the main text.

The internal standards 4-phenylbutyric acid and malonic acid as well as the anhydrous Na₂SO₄ were purchased from BDH. Ethylacetate and diethylether were obtained from Merck Chemicals. The derivatization reagents O-bis(trimethylsilyl)trifluoroacetamide (BSTFA), trimethylchlorosilane (TMCS) and pyridine were purchased from Sigma Chemical Company.

1.2 GC–MS analysis

An Agilent GC–MS system was used in this study and consisted of a model 7890A gas chromatograph, a model 5975C mass selective detector, a Hewlett Packard 5970C quadrupole mass spectrometer and Agilent Chemstation. GC separation was achieved on a WCOT fused silica capillary column [30 m x 0.32 mm (i.d.)] coated with SE30 CB, film thickness 0.30 µm (Machery Nagel).

Samples were introduced into the column via a splitless injector at 280°C. The initial oven temperature was kept at 60°C for 2 minutes and then programmed to rise 6°C/min to a final temperature of 280°C. This temperature was maintained for 5 minutes. Helium was used as carrier gas at a pressure of 60 kPa and at a constant flow rate of 1 ml/min. The column was inserted directly into the ion source with an interface temperature of 280°C. The mass spectra of all GC peaks were generated by a mass spectrometer operated in the electron impact (EI) mode, with electron energy of 70 eV. The MS source and quadrupole temperatures were 250°C and 150°C respectively.

1.3 Qualitative analysis of metabolites in QC samples (metabolite identification)

After GC separation, each peak at an indicated RT was analysed in EI mode in order to investigate the fragmentation pattern of each compound. The m/z values selected for feature identification were either the base peak ion or one of the more abundant characteristic fragment ions.

Deconvolution and data analyses were conducted using AMDIS software (Version 2.71) linked to NIST Mass Spectral Search Program for the NIST/EPA/NIH Mass Spectral Library (Version 2.0F, built Oct. 8, 2008). Where authentic standards were available, their respective response factors were used, and for those compounds where no authentic standards were available, a response factor of 1 was assumed. The analytical setting of the AMDIS software was as follows: minimum factor – 60%, threshold – “off”, Scan Direction – “high to low”, and type of analysis – “Use of an internal standard for RI”. The deconvolution settings were: component width – 12, adjacent peak subtraction – 1, resolution – medium, sensitivity – low, and shape requirements – low. The first hit of identified compounds and integrated area of the peaks were exported to Microsoft Excel.

2. Data used for the alcohol intervention study

Supplementary Table S1 shows an excerpt of the data used for the alcohol intervention study (the full data set is given as a separate file in Excel format).

Supplementary Table S1

An excerpt of the data used for the alcohol intervention study

		1	2	3	4	5	6	7	8	9	...	119	120
Batch	Time	Glyoxylic-1,2-Dihyd	1,2-Dihyd	Lactic-acid	Hexanoic-2-Hydroxy	Glycolic-a	Oxalic-acid	2-Hydroxy				Octadecar	Eicosanoic
1	0	0.181	0.825	6.424	8.188	0.198	8.959	7.566	2.127	3.362		1.918	0.030
1	1	0.004	0.863	0.586	6.011	0.194	7.354	3.867	1.043	1.983		1.086	0.035
1	2	0.000	3.348	23.576	73.513	1.017	37.804	11.328	1.472	5.640		0.125	0.014
1	3	0.004	1.715	2.798	14.757	0.219	11.194	1.895	1.719	2.231		5.534	0.057
1	4	0.004	1.544	3.411	16.042	0.525	12.337	3.209	0.507	1.175		3.472	0.012
2	0	0.380	6.090	3.299	10.635	0.213	13.432	5.542	1.656	2.546		1.380	0.006
2	1	0.577	6.087	5.495	31.238	0.207	51.693	12.475	3.070	8.968		7.982	0.015
2	2	2.431	10.583	6.763	19.799	0.040	24.936	3.507	4.564	4.293		16.834	0.064
2	3	0.165	6.804	2.695	8.701	0.031	13.519	1.426	2.046	2.385		3.687	0.016
2	4	0.392	5.628	0.717	11.225	0.016	22.899	4.268	1.535	3.195		2.861	0.018
3	0	0.057	6.082	1.117	5.128	0.017	7.560	2.057	0.249	1.832		0.587	0.002
3	1	0.263	6.834	1.312	6.123	0.005	11.984	2.682	0.708	2.046		0.579	0.008
3	2	0.074	7.555	3.482	11.516	0.041	10.888	1.091	0.437	0.051		4.187	0.009
3	3	0.009	7.164	0.762	2.578	0.011	4.549	0.295	0.566	0.173		1.701	0.000
3	4	0.142	7.788	2.470	7.058	0.023	10.243	1.352	2.480	0.027		5.931	0.012
4	0	1.260	5.450	14.115	13.301	0.008	19.194	3.939	1.727	3.589		1.668	0.004
4	1	0.786	6.998	4.797	7.539	0.005	11.048	1.319	0.234	0.600		2.648	0.019
4	2	5.384	8.793	7.255	22.786	0.074	34.901	1.477	1.401	2.113		20.830	1.148
4	3	0.452	6.635	2.458	3.065	0.002	6.041	0.249	0.109	0.308		1.754	0.022
4	4	10.047	5.740	14.662	20.996	0.176	32.883	3.912	0.401	2.180		5.673	0.038
5	0	0.190	5.314	1.160	13.823	0.059	18.080	6.760	1.977	3.394		1.605	0.040
5	1	0.348	5.662	2.762	16.770	0.036	20.638	8.213	2.311	4.511		3.696	0.830
5	2	0.004	13.903	16.727	58.947	0.143	44.790	9.347	11.309	9.393		20.223	0.187
5	3	0.041	9.541	9.511	52.548	0.150	29.792	5.183	1.630	3.260		21.992	0.056
5	4	3.452	6.023	1.227	12.596	0.041	19.668	4.064	1.310	1.860		5.391	0.045

3. Details of the 172 identified features

Supplementary Table S2 gives the details of each of the 172 identified and classified variables. A reference value for the normal concentration range in urine for adults of each identified variable is also given.

Supplementary Table S2 Identification, classification and reference ranges of 172 identified variables

	Tentative variable name based on AMDIS and the library used	RT (min)	Variable no.	Index reference	Reference range (μmol/mmol Cr)	Excluded variable and criteria
1	Glyoxylic acid	7.769	1	HMDB00119	1.27 (0.46–3.26)	
2	2-Undecenoic acid	8.029		CS 21171839	n/r	Not a recognized biological substance
3	1,2-Dihydroxyethane	8.693	2	KEGG C01380	n/r	
4	1,2-Dihydroxypropane	9.211	3	ChEBI 16997	n/r	
5	Boric acid	9.24		HMDB35731	n/r	Not a recognized biological substance
6	Phenol	10.158		HMDB00228	4.8 (0.6–12.8)	Not a recognized biological substance
7	O-acetylphenol	10.323		HMDB32568	n/r	Not a recognized biological substance
8	Lactic acid	10.652	4	HMDB00190	3.9–9.8	
9	Hexanoic acid	10.75	5	HMDB00535	n/q	
10	2-Hydroxyisobutyric acid	10.881	6	HMDB00729	4.4–7.6	
11	Glycolic acid	1.952	7	HMDB00115	64.0 (21.0–107.0)	
12	Oxalic acid	12.229	8	HMDB02329	27.00 (0.00–54.00)	
13	2-Hydroxybutyric acid	12.549	9	HMDB00008	2.8 (1.2–6.9)	
14	4-Ketovaleric acid	12.603	10	HMDB00720	1.3 (0.4–2.0)	
15	4-Cresol	12.834		HMDB01858	46.0 (1.2–118.9)	Questionable biological substance/origin
16	3-Hydroxypropionic acid	12.882	11	HMDB00700	8.1 (3.1–11.8)	
17	2-Methyl-2-hydroxybutyric acid	12.921	12	HMDB01987	0.8 (0.4–1.5)	
18	Sulphate	12.954		HMDB01448	2407.89 (2171.05–2697.36)	Not a recognized biological substance
19	2-Hexenoic acid	13.057	13	HMDB10719	n/r	
20	3-Hydroxybutyric acid	13.421	14	HMDB00357	1.4–2.2	
21	3-Hydroxyisobutyric acid	13.442	15	HMDB00023	11 (4.1–19)	
22	2-Hydroxyisovaleric acid	13.665	16	HMDB00407	0.23–0.42	
23	N-Crotonylglycine	13.82	17	CAS 71428-89-2	n/r	Glycine conjugation product of crotonic acid
IS	<i>Malonic acid</i>	14.402	IS-2			
24	2-Methyl-3-hydroxybutyric acid	14.518	18	HMDB00354	4.2 (1.6–6.7)	
25	Methylmalonic acid	14.74	19	HMDB00202	1.8 (0.00–3.6)	
26	3-Hydroxyisovaleric acid	14.802	20	HMDB00754	11 (6.9–25)	
27	Benzoic acid	15.208		HMDB01870	4.2 (1.9–6.5)	Preservative of the vehicle
28	2-Ethyl-3-hydroxypropionic acid (2-ethylhydracrylic acid)	15.277	21	HMDB00396	2.1 (1.3–2.9)	
29	3-Hydroxyvaleric acid	15.462	22	HMDB00531	1 (0–2)	
30	Acetoacetic acid	15.532	23	HMDB00060	0.15 (0.01–0.58)	
31	2-Methyl-3-ketobutyric acid	15.601	24	HMDB03771	1.0 (0.0–2.0)	

	Tentative variable name based on AMDIS and the library used	RT (min)	Variable no.	Index reference	Reference range (μmol/mmol Cr)	Excluded variable and criteria
32	4-Pyridinecarboxylic acid	15.741		HMDB60665	n/r	Derivatization artifact
33	Octanoic acid	15.899	25	HMDB00482	n/q	
34	2,2-Dihydroxyacetic acid	15.929		CS 2005843	n/r	Not a recognized biological substance
35	2-Ketoisovaleric acid	15.6	26	HMDB00019	0.13 (0.00–0.54)	
36	N-Acetylglycine	15.66	27	HMDB00532	n/r	
37	Phosphoric acid	16.288	28	HMDB02142	784 (425–1170)	
38	Octanoic acid	15.899	29	HMDB00482	n/q	
39	3-Ketovaleric acid	16.198	30	CS 388751	n/r	Can be of endogenous origin in certain metabolic disorders
40	Phenylacetic acid	16.37	31	HMDB00209	0.3 / 1.9	
41	Ethylmalonic acid	16.379	32	HMDB00622	1.16–2.99	
42	Glycerol	16.687	33	HMDB00131	13 (4–19)	
43	2-Octenoic acid	16.954	34	HMDB00392	n/r	
44	Succinic acid	17.024	35	HMDB00254	9.9 +/- 5.0	
45	2-Hydroxyglutarylactone	17.141		n/a	n/r	From 2-hydroxyglutaric acid (see no. 104)
46	3-Hydroxycaproic acid	17.169	36	HMDB02203	n/r	
47	1,2-Dihydroxybenzene	17.263	37	HMDB00957	4.06 +/- 1.80	
48	Methylsuccinic acid	17.382	38	HMDB01844	2.3 (0.8–10.8)	
49	Pyrimidinedione	17.48		n/a	n/a	Derivatization artifact
50	Uracil	17.555	39	HMDB00300	9.5 (2.6–22.8)	
51	5-Hydroxyvaleric acid	17.52	40	HMDB61927	n/r	
52	Citraconic acid	17.81		HMDB00634	1.6 (0.9–2.1)	Not a recognized biological substance
53	Glyceric acid	17.815	41	HMDB00139	1.7 (0.2–6.0)	
54	Glutaconic acid	17.839	42	HMDB00620	3.1 (1.2–3.1)	
55	Fumaric acid	17.889	43	HMDB00134	0.75–1.2	
56	Nonanoic acid	18.115	44	HMDB00847	n/r	
57	5-Hydroxyhexanoic acid	18.224	45	HMDB00525	2.7 (0.8–5.7)	
58	N-Isobutyrylglycine	18.297	46	HMDB00730	n/q	
59	2,3-Dihydroxybutanoic acid	18.391	47	HMDB00498	5.0 +/- 3.0	
60	Lactylactate	18.762		CS 91775	n/r	From lactic acid
61	1,3-Dihydroxybenzene	18.62		HMDB32037	n/r	Questionable biological substance
62	Hydroxymalonic acid	18.913		HMDB35227	n/r	Derivative from IS-2
63	Glutaric acid	18.975	48	HMDB00661	1.3 (0.6–2.6)	
64	Phenoxyacetic acid	19.083		HMDB31609	n/r	Questionable biological substance
65	Benzamide	19.153		HMDB04461	n/r	Not a recognized biological substance
66	2,3,4-Trihydroxybutyric acid-lactone	19.32		n/a	n/r	From 2,3,4-trihydroxybutyric acid
67	N-Butyrylglycine	19.46	49	HMDB00808	0.24 +/- 0.36 [infant]	
68	3-Methylglutaric acid	19.442	50	HMDB00752	1.0–6.5	
69	N-Isobutyrylglycine(2)	19.46		HMDB00730	n/q	Second peak (see no. 58)
70	Malonic acid-tri-TMS	19.602		Agilent G1676AA	n/r	Derivative of IS-2
IS	4-phenylbutyric acid	19.636	IS-1			
71	3-Methylglutaconic acid	19.698	51	HMDB00522	6.2 (2.8–8.3)	
72	3-Hydroxyadipylactone	19.748	52	n/a	n/r	From adipic acid (see no. 86)
73	N-Acetyl-leucine	19.768	53	HMDB11756	n/r	

	Tentative variable name based on AMDIS and the library used	RT (min)	Variable no.	Index reference	Reference range (μmol/mmol Cr)	Excluded variable and criteria
74	3,4-Dihydroxybutyric acid	19.952		HMDB06118	34.9 (12.5–57.2)	Questionable biological substance
75	Decanoic acid	20.1	54	HMDB00511	n/q	
76	N-Isovalerylglycine	20.22	55	HMDB00678	2.0 (0.4–4.0)	
77	Parabanic acid	20.308		CID 67126	n/r	Medication/Artifact
78	2,2-Dihydroxymalonic acid	20.473		CS 61695	n/r	Derivative from IS-2
79	3-Ethylglutaric acid	20.54	56	n/a	n/a	Potential ethanol derivative
80	5-Methyldihydropyrimidine-2,4(1H,3H)-dione	20.572		CS 84456	n/r	Not a recognized biological substance
81	3,5-Dihydroxy-3-methylvaleric acid	20.697	57	HMDB00227	0.14 (0.06–0.22)	
82	N-Isovalerylglycine(2)	20.823		HMDB00678	2.0 (0.4–4.0)	Second peak (see no. 76)
83	Pyroglutamic acid	21.128	58	HMDB00267	14.0 +/- 7.2	
84	3-Methylglutaconic acid	20.626	59	HMDB00522	6.2 (2.8–8.3)	Second peak (see no. 71)
85	Citramalic acid	20.717		HMDB00426	2.4 (1.0–4.8) [NMR analysis]	Questionable biological substance
86	Adipic acid	20.94	60	HMDB00448	5.1 (0.8–35)	Can be of endogenous origin in certain metabolic disorders
87	Malic acid	20.942	61	HMDB00156	2.0 (0.7–5.3)	
88	2-Deoxy-3,5-dihydroxypentonic acid-g-lactone	20.971	62	n/a	n/a	
89	2-Piperidinecarboxylic acid	21.14		HMDB00070	0.28 (0.0–0.56)	Questionable biological origin
90	4-Hydroxycyclohexylcarboxylic acid	21.243		HMDB01988	6.1 (1.1–28.1) [Adolescent]	Questionable biological substance
91	5-Hydroxyhydantoin	21.323		CID 4157426	n/r	Derivatization artifact
92	N-Acetylthreonine	21.356	63	CS 3841173	n/r	
93	3-Methyladipic acid	21.518	64	HMDB00555	2.9 (0.7–10.5)	
94	5-(Hydroxymethyl)furan-2-carboxylic acid	21.651		HMDB02432	1.7	Questionable biological substance
95	2,4-Hexadienedioic acid	21.65		HMDB29581	n/r	Questionable biological substance
96	2-Hydroxyphenylacetic acid	22.081	65	HMDB00669	2.0 (0.9–4.5)	
97	N-Tiglylglycine	21.736	66	HMDB00959	0.78–1.2	
98	Threitol	21.744	67	HMDB04136	19.3 (5.3–32.7)	
99	2-Hydroxy-2-methylglutaric acid	21.901	68	Yancey <i>et al.</i> , 1986	n/r	
100	2-Hydroxyphenylacetic acid	22.081	69	HMDB00669	2.0 (0.9–4.5)	
101	N-3-Methylcrotonylglycine	22.128	70	HMDB00459	1.0 (0.0–2.0)	
102	2,3,4-Trihydroxybutyric acid	22.275		HMDB00613	20.0 +/- 8.0	Questionable biological substance
103	N-2-Pentenylglycine	22.358	71	n/a	n/a	
104	2-Hydroxyglutaric acid	22.453	72	HMDB00694	0.8–52	
105	3-Hydroxyglutaric acid	22.49	73	HMDB00428	<11.51 [children]	
106	2,3,4-Trihydroxybutyric acid(2)	22.52	74	HMDB00613	20.0 +/- 8.0	Second peak (see no. 102)
107	2,4,6-Trihydroxypyrimidine	22.665		HMDB41833	n/r	Derivatization artifact
108	Pimelic acid	22.678	75	HMDB00857	2.2 (0.7–4.0)	
109	Hexahydropyrimidine-2,4,5-trione	22.696		n/a	n/a	Not a recognized biological substance
110	3-Ketoglutaric acid	22.841	76	HMDB13701	0–0.11	
111	N-Hexanoylglycine	22.869	77	HMDB00701	1.00 (0.0–2.0)	
112	2-Ketoglutaric acid	22.949	78	HMDB00208	2.87 (0.18–14.3)	
113	3-Hydroxyphenylacetic acid	22.753	79	HMDB00440	0.6 (0.4–0.9)	
114	3-Hydroxy-3-methylglutaric acid	22.992	80	HMDB00355	3.2 (1.1–5.2)	

	Tentative variable name based on AMDIS and the library used	RT (min)	Variable no.	Index reference	Reference range (μmol/mmol Cr)	Excluded variable and criteria
115	4-Hydroxycyclohexylacetic acid	23.047		HMDB00909 (trans)	n/r	Not a recognized biological substance. Alternative reference: CS 13628091 (cis)
116	4-Hydroxybenzoic acid	23.143	81	HMDB00500	1.8 (0.7–29)	
117	4-Hydroxyphenylacetic acid	23.249	82	HMDB00020	6.0 (2.4–9.7)	
118	2,5-Furandicarboxylic acid	23.255		HMDB04812	1.9 (0.1–5.4) [Adolescent]	Derivatives found in food prepared by strong heating. Not a recognized biological substance
119	N-Acetylaspartic acid	22.18	83	HMDB00812	4.66 +/- 1.14	
120	2,3,4,5-Tetrahydroxypentanoic acid-1,4-lactone	23.496	84	CAS 179091-67-9	n/r	From ascorbic acid (see no. 151)
121	Dodecanoic acid	23.807	85	HMDB00638	0.03 +/- 0.02	
122	2-Furanylcarbonylglycine	23.639		HMDB00439	9.95 (2.00–18.66)	Not a recognized biological substance (see no. 118)
123	N-Cyclohexylsulfamic acid	23.849		HMDB31340	n/r	Not a recognized biological substance
124	N-Acetylthranilic acid	23.908	86	CID 6971	n/r	
125	Octenedioic acid	23.897	87	HMDB00341	n/r	
126	Isocitric acid-lactone	24.048		CID 98259	n/r	From isocitric acid (see no. 145)
127	Suberic acid	24.255	88	HMDB00837	0.5 (0.0–2.9)	
128	Erythro-pentonic acid	24.561	89	n/a	n/r	Possible CAS 74742-30-6
129	Acetyl-4-phenol	24.595		CS 7189	n/r	Not a recognized biological substance
130	Tricarballic acid	25.195		HMDB31193	n/r	Derivatization artifact
131	Citric acid-diethylester	25.195	90	CAS 19958-02-2	n/r	
132	Aconitic acid	25.195	91	HMDB00461	13 (2.7–44)	
133	Vanillic acid	25.283	92	HMDB00484	1.0 (0.0–2.5)	
134	Homovanillic acid	25.347	93	HMDB00118	5.6 (2.1–47.3)	
135	Pyrrole-2-carboxylic acid	25.551	94	HMDB03094	0.03 +/- 0.02	
136	2,5-Dihydroxybenzoic acid	25.686	95	HMDB00152	1.35 +/- 0.79	
137	4-Hydroxymandelic acid	25.737	96	HMDB00822	0.98–1.5	
138	Azelaic acid	25.819		HMDB00784	4.8 (1.3–15)	Questionable biological substance
139	Hippuric acid-diTMS	26.055		CS 454967	n/r	Derivative of hippuric acid
140	Citric acid-monoethylester	25.923	97	CID 57345948	n/r	
141	Hippuric acid-TMS	26.058	98	HMDB00714	27.92–932.66	
142	3,5-Dihydroxybenzoic acid	26.281	99	HMDB13677	0.218 +/- 0.124	
143	3,4-Dihydroxybenzoic acid	26.34		HMDB01856	0.048 +/- 0.008	Questionable biological substance
144	3,4-Dihydroxyphenylacetic acid	26.454	100	HMDB01336	0.9 (0.6–1.3)	
145	Isocitric acid	26.552	101	HMDB00193	56.8 (19.4–119.1)	
146	Citric acid	26.584	102	HMDB00094	172.8 +/- 87.4	
147	Tetradecanoic acid	26.65	103	HMDB00806	0.06 +/- 0.03	
148	3-Hydroxyphenylhydracrylic acid	26.844		HMDB02643	5.9 (1.4–22.1)	Questionable biological substance
149	Methylcitric acid	27.008	104	HMDB03610	0.59 (0.05–1.15)	
150	Vanillylmandelic acid	27.168	105	HMDB00291	1.1–1.7	
151	Ascorbic acid	27.303		HMDB00044	32.5 (4.6–78)	Vehicle additive
152	Hydantoinpropionic acid	27.255		HMDB01212	n/r	Questionable biological substance

	Tentative variable name based on AMDIS and the library used	RT (min)	Variable no.	Index reference	Reference range ($\mu\text{mol}/\text{mmol Cr}$)	Excluded variable and criteria
153	4-Hydroxyphenyllactic acid	27.571	106	HMDB00755	1.1 (0.2–2.6)	
154	1H-Indole-3-acetic acid	28.028	107	HMDB00197	1.5–2.6	
155	3,4-Dihydroxyphenylpropionic acid	28.131		HMDB00423	0.27 +/- 0.079	Questionable biological substance
156	3-(4-Hydroxy-2,5-dioxoimidazolidin-4-yl) propanoic acid	28.155		n/a	n/a	Not a recognized biological substance
157	3-Methoxy-4-hydroxyphenylhydracrylic acid	28.761		Smith, 1969	n/r	Not a recognized biological substance
158	2,4,6-Trihydroxybenzoic acid	28.771		CID 66520	n/r	Not a recognized biological substance
159	Palmitic acid	29.416	108	HMDB00220	11 (6.0–23)	
160	2-Hydroxyhippuric acid	29.501	109	HMDB00840	0.5	
161	Glucuronic acid	29.734	110	HMDB00127	9.7 (3.7–20.6)	
162	4-Methoxy-3-hydroxycinnamic acid (Isoferulic acid)	29.886	111	HMDB00955	0.394 +/- 0.075	
163	N-Acetyltyrosine	29.886	112	HMDB00866	<10 [children]	
164	4-Hydroxyphenylpyruvic acid	29.738	113	HMDB00707	1.65 (0.15–8.74)	
165	Heptadecanoic acid	30.693	114	HMDB02259	n/r	
166	N-Cinnamoylglycine	30.723	115	HMDB11621	n/r	
167	4-Hydroxyhippuric acid	31.15	116	HMDB13678	0–14	
168	Phenylacetylglutamine	30.71	117	HMDB06344	47.03 (3.84–85.51)	
169	4-Hydroxyhippuric acid(2)	31.423		HMDB13678	0–14	Second peak (see no. 167)
170	Oleic acid	31.583	118	HMDB00207	5.2 (0.3–13)	
171	Octadecanoic acid	31.904	119	HMDB00827	2.9 (1.6–6.6)	
172	Eicosanoic Acid	33.85	120	HMDB02212	n/r	

The original list of 172 variables contained substances that were excluded from the metabolomics analysis to assess the influence of acute alcohol consumption. Only the 120 included substances are numbered in column 4 of the table. The reasons for the exclusion of the other 52 substances are given in the final column of the table.

Categories for exclusion of variables were:

1. Exogenous contaminants that are not recognized biological substances.
2. Substances of a questionable biological function or origin, based on literature assessments.
3. Artifacts formed from chemicals (e.g. urea) present in the urine.
4. Artifacts due to the reaction conditions (e.g. formation of lactones, used for correction of the parent substance).
5. Artifacts due to formation of an additional TMS derivative (used for correction of the parent substance).
6. Substances showing multiple peaks in the GC-profile (used for correction to only one substance).

4. Statistical Methods & Additional Results

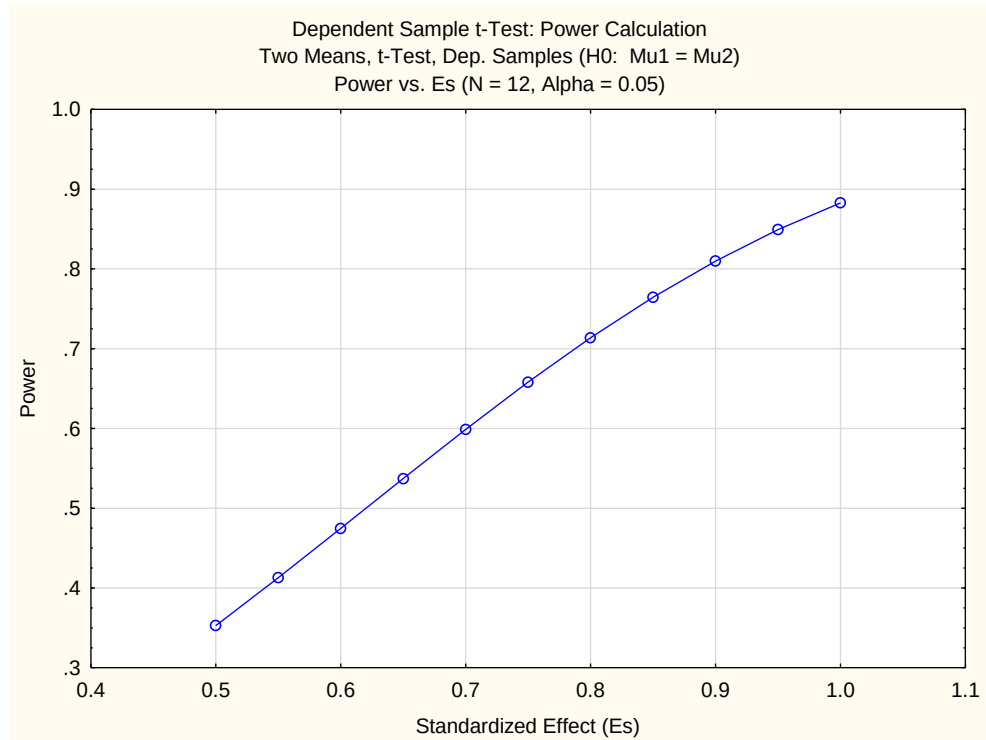
4.1 Power

Only a limited number of subjects could be included in an intervention study of this nature, i.e. exposure to a substance (alcohol) with known health risks. It was therefore necessary to ensure that the small sample size would not prevent the quantification of the effect of the intervention. A power calculation was performed assuming the following:

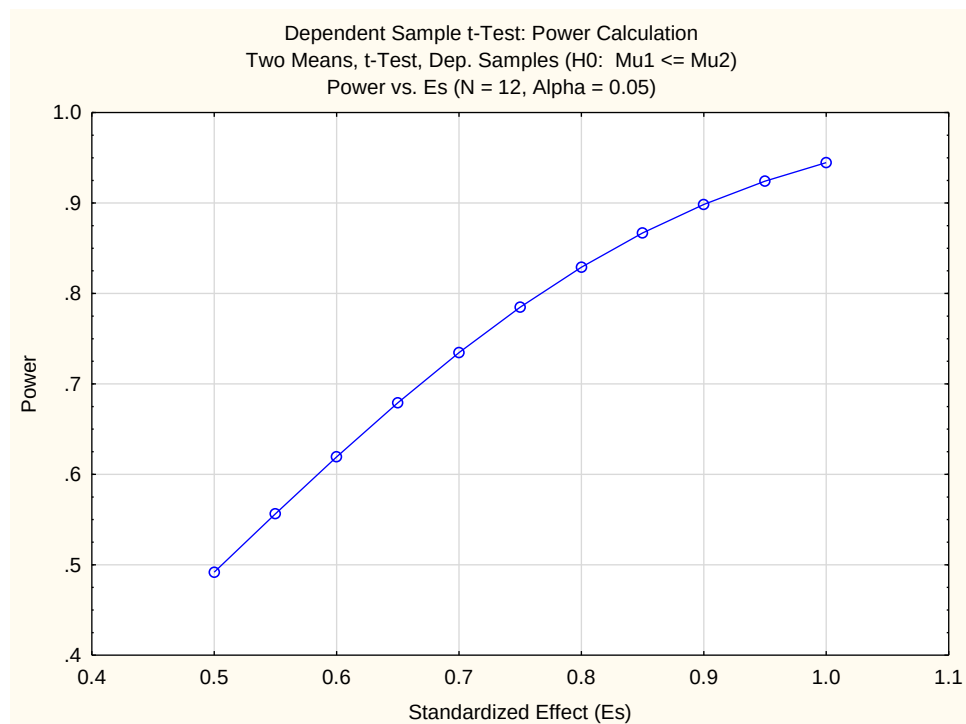
- (i) The effect of the intervention would be large, with an expected effect size of at least 0.8 when calculated based on Cohen's d-value¹. This was considered a valid assumption given the design of the study, i.e. consuming a relatively large volume of alcohol on an empty stomach.
- (ii) The dependence between samples from the same individual would be taken into account by using the dependent samples, or paired t-test, to assess the significance of the effect of the intervention. Though this method is a parametric univariate method comparing only two groups, it was deemed appropriate as we could transform the data to achieve normality, we did not know beforehand how many metabolites would be extracted, and the most dramatic effect of the intervention was expected to be between time 0 (before consumption of the alcohol) and 1 hour after consumption based on a pilot study.
- (iii) A significance level of 5% for a two-tailed test and a sample size of 12. We did not want to speculate on the metabolites that would be extracted and hence could not speculate on the direction of the expected change. The two-tailed hypothesis was therefore selected.

The power under these conditions is plotted against the effect size in Supplementary Fig. S1, from which it is evident that power values above 0.7 are achieved for effect sizes of 0.8 or higher. This was acceptable in the current context. The power curve given a one-sided hypothesis is also included for reference purposes to show the increased power.

a



b



Supplementary Figure S1. Power curves

Power curves for a dependent samples t-test for a sample size of 12 and a 5% significance level, when setting a two-sided (a) and one-sided (b) alternative hypothesis.

4.2 Data pre-processing

The collection and analysis of biological samples culminated in 12 subjects observed across 5 time points and measured over 120 metabolites. A 50% zero filter was applied, taking time into account — that is, variables were excluded if they contained more than 50% zero-values in all 5 times. Two variables were removed based on these criteria. The remaining zero-values were imputed for each variable by randomly generating numbers from the tail of beta-distribution fitted to the non-zero observations. The resulting random numbers were all smaller than half of the minimum observed value. The data were then log transformed and centred prior to further statistical analysis.

4.3 Statistical methods for variable selection

It was decided to deconstruct the three-dimensional data tensor into less complex cross-sections in order to select a subset of statistically important metabolites to aid biological interpretation of the observed perturbations. The assumption was made that the most biologically acute effect of the intervention would peak one hour after alcohol consumption. The cross-section between time 0 (before consumption of the alcohol) and one hour after consumption was therefore analysed using univariate and multivariate statistical methods.

The univariate nonparametric Wilcoxon signed-rank test was used to assign significance levels (p -values) to changes in metabolite levels from time 0 to time 1. The dependent t-test was also applied, after data transformation to ensure normality, but not used as current research revealed that the Wilcoxon signed-rank test to be a suitable alternative with marginally more power for a non-normally distributed data set of smaller sample size².

Multivariate partial least squares-discriminant analysis (PLS-DA) is a method that constructs a regression model to predict group membership by projecting variance in the metabolites measured, as well as in the observed group membership, to new spaces. PLS-DA models can easily overfit and produce models with inflated predictive ability unless extensively validated through test data and permutation testing. For this reason, the PLS-DA results are used cautiously as the small sample size did not allow for proper validation. The results used are limited to the VIP values, which are produced by the PLS-DA model for each metabolite as an indication of its predictive strength. Given our concern here, it is again important to emphasize that the aim of this selection was not to model the observed data, but rather to gain a deeper understanding of the predominant biological changes occurring due to the intervention. Ranking and selecting metabolites according to VIP values enabled us to identify metabolite combinations which dominated the observed change in metabolomic state. Ultimately, we revert back to the raw data (means and standard deviations) to interpret the resulting list against established metabolic pathways, and including proposals for extension or modification of them, resulting in a representation of a global metabolite profile reflecting the metabolic consequences of the alcohol consumption.

PCA is similar to PLS-DA in that observed data are also projected, but to new spaces that maximize variation along fewer hyperplanes, not taking the group membership into consideration.

A subset of metabolites was subsequently identified to gain a deeper understanding of the predominant biological changes occurring due to the intervention. The selection was not made for the purpose of modelling the effect of the intervention in time nor to predict group

membership. Changes in metabolites levels were then ranked based on their multivariate VIP values. The significance of changes in high-ranking variables was then established through fold change (FC) values and non-parametric Wilcoxon signed-rank test (WRT) p -values. The selection criteria were then as follows: $VIP \geq 1.0$, $WRT\ p \leq 0.05$ and $|FC| \geq 1.5$.

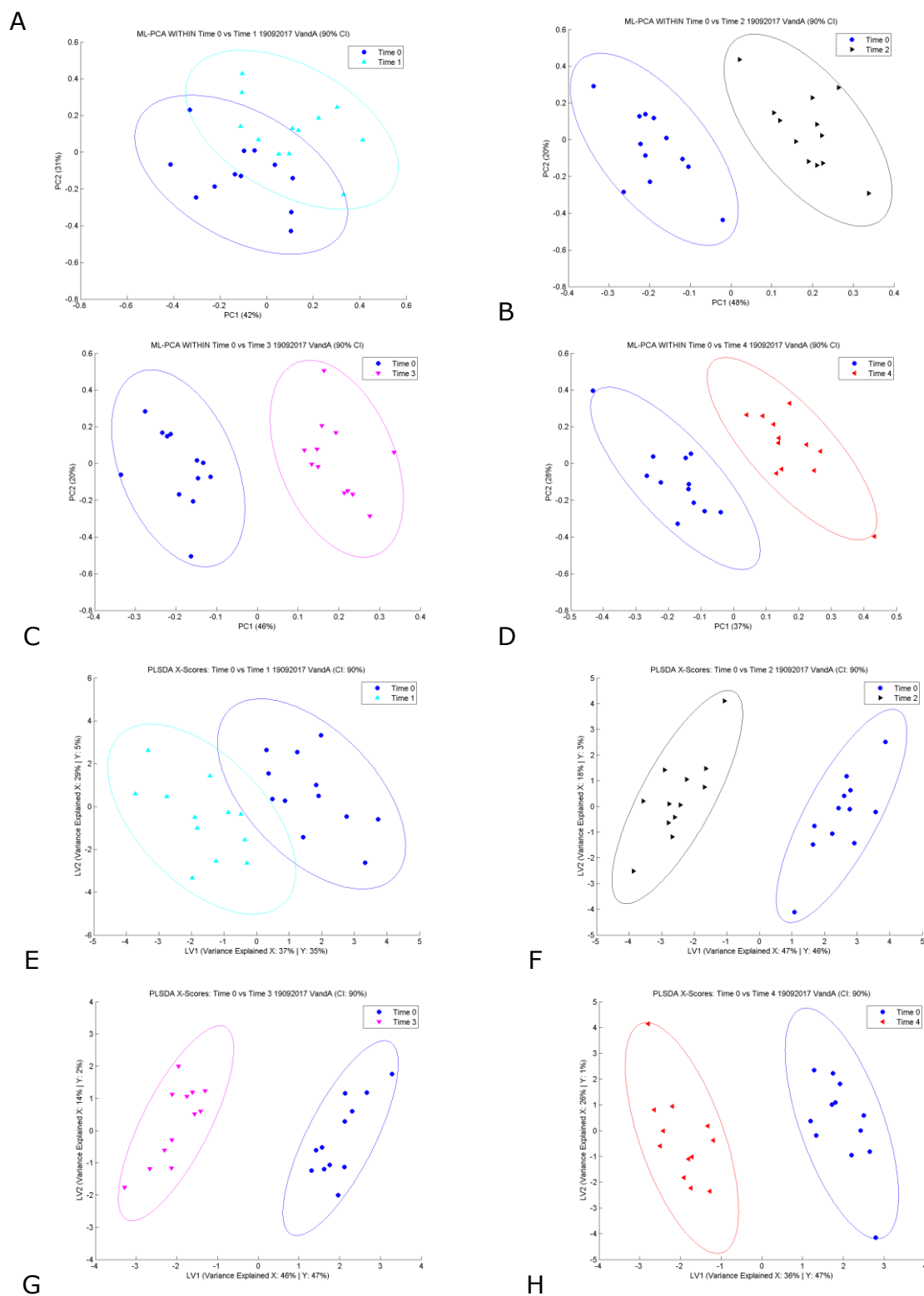
4.4 Correlation analysis

Associations between metabolites shortlisted for biological interpretation were assessed using Spearman's rho or rank order correlation coefficient. This method is nonparametric as this takes the ranks of the data as inputs. The correlation coefficients (r-values) produced are seen as biologically relevant if $|r| \geq 0.5^1$.

4.5 Statistical methods for repeated measures data

The data tensor evaluated here has a very specifically designed structure with the same individuals being measured over and over, hence a repeated measures design over 5 points in time. Although the Wilcoxon signed-rank test accounts for groups of data not being independent but paired, standard PCA and PLS-DA applications do not. To understand the importance of this information for these models, a multi-level PCA and PLS-DA on the within-subject variation was also performed.

Supplementary Fig. S2 provides the scores plots corresponding to the cross-sections of the data used in the main text. The PCA scores plots are comparable with those included in the main text, showing some differentiation one hour after the alcohol consumption (Fig. S2A), followed by complete separation after 2 and 3 hours. The separation is, however, retained after 4 hours in the multi-level PC model. That said, a slight yet progressive return after 3 and 4 hours to the profile at time 0 is still evident when considering the decreasing variance explained by the first principal component. The PLS-DA plots again showed a complete separation for all four times following alcohol consumption relative to time 0.



Supplementary Figure S2. PCA and PLS-DA plots following alcohol consumption. Input data were from the 120 quantified metabolites for time 0 vs time 1 (A & E), time 0 vs time 2 (B & F), time 0 vs time 3 (C & G), and time 0 vs time 4 (D & H) following alcohol intake. The samples for time 0 were collected just prior to the interventions and are therefore regarded as the control samples.

Finally, the entire data tensor was modelled using PCA. This was achieved by unfolding the data along the time dimension³.

This approach allows for the projection of data to fewer dimensions, while retaining the maximum amount of variation and understanding of the changes observed in time. This can then be observed and compared over all individuals, by averaging scores, or for each individual to allow for comparison. In addition, the contribution of each metabolite to the projection can be established based on their loadings, but more noteworthy, based on the directional loadings vectors as reflected in the bi-plots (Fig. 4 in the main text).

4.6 The vehicle/hippuric acid effect

The consumption of the flavoured water as vehicle is an intervention in its own right, given the relatively high concentration of sodium benzoate, used as a preservative. We have previously described the effect of the vehicle consumption⁴. From this investigation it became clear that hippuric acid dominates the urine profile, which we took in consideration when using flavoured water as vehicle for alcohol consumption — we applied a paired PLS–DA on the original data, excluding the information on hippuric acid, and compared the VIP values from this analysis to those from the unpaired PLS–DA which included hippuric acid. Both analyses produced 27 metabolites with a VIP > 1.0, with only the 12 metabolites defined as important indicators, and listed in Table 1, being common to both analyses when applying the additional criteria of WRT $p \leq 0.05$ and $|FC| \geq 1.5$. These results imply that the presence of hippuric acid in the data set did not influence the metabolites listed in Table 1, and, per implication, did not affect the outcome of the study.

4.7 Statistical software

Supplementary Table S3 lists the data analysis software used to perform the different statistical analyses discussed throughout this section.

Supplementary Table S3 List of statistical software

Analysis	Software
Univariate statistics	MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, MA, USA.
Multivariate statistics	PLS_Toolbox 8.2.1 (2016). Eigenvector Research, Inc., Manson, WA, USA 98831; software available at http://www.eigenvector.com .
Correlation analysis	T. Wei and V. Simko (2016). corrplot: Visualization of a Correlation Matrix. R package version 0.77. https://CRAN.R-project.org/package=corrplot
Power curves	Dell Inc. (2016). Dell Statistica (data analysis software system), version 13. software.dell.com .
ML–PCA	Matlab m files created by Biosystems Data Analysis Group, Universiteit van Amsterdam. www.bdagroup.nl/content/Downloads/software . Copyright 2008
ASCA	Matlab using the statistics toolbox and code provided by Gooitzen Zwanenburg (available under APACHE Licence 2.0 http://www.apache.org/licenses/LICENSE-2.0.html).

5. Example of Informed Consent Form

INFORMED CONSENT TO PARTICIPATE IN EXPERIMENT WITH ETHICAL APPROVAL FROM THE NORTH-WEST UNIVERSITY

TITLE:

An investigation into the metabolic responses to acute alcohol consumption

AIM OF EXPERIMENT:

To determine the metabolic perturbations in young males in a fasted state, due to the consumption of a fixed acute dose of alcohol and/or NAD, as investigated by a metabolomics methodology.

INVESTIGATOR'S NAME: Cindy Irwin

SUPERVISOR: Prof. C.J. Reinecke

INVESTIGATOR SITE NAME & ADDRESS: Centre for Human Metabonomics
NWU, Potchefstroom Campus
Private Bag X6001
POTCHEFSTROOM 2520
South Africa
Tel: 018 299 2309
Fax: 018 293 5248

INTRODUCTION

The North-West University's Centre for Human Metabolomics aims to investigate perturbations associated with human metabolism by means of a metabolomics approach. Metabolomics is a biochemical technique, involving a comprehensive study of low molecular weight biomolecules, commonly known as metabolites. Biochemical analysis of biological samples (such as urine or blood) provides a large comprehensive list of metabolites. The data are analysed by means of bioinformatics, a field of science incorporating statistical multivariate analysis, providing information used to determine/distinguish any potential irregularities within the metabolic profile. The focus is the identification of very specific metabolites that can statistically discriminate between normal and abnormal metabolic situations. These metabolites are known as biomarkers.

PURPOSE OF THE EXPERIMENT

This experiment constitutes part of the investigator's M.Sc. thesis involving the study into the metabolic perturbations associated with alcohol consumption. This experiment is aimed at determining the changes in the metabolite profile that occur after the consumption of an acute alcohol dose and what, if any, effect the administration of NAD together with the alcohol has to prevent these metabolic changes. To minimize variation, a homogeneous, defined experimental group, namely young males between the ages of 20 and 30 years in an overnight fasted state, will be used. The biological specimens used in this experiment will be a blood sample taken before the start of the experiment, as well as urine samples taken at defined intervals of time, followed by a metabolomics analysis.

PROTOCOL OF EXPERIMENT

Acute Alcohol and/or NAD Dose Study:

- Participants are required to abstain from alcohol consumption at least 48 hours preceding this experiment; as well as abstaining from consumption of food prior to initiation of the experiment

- (i.e. overnight fasted state).
- A questionnaire involving the basic clinical profile, including history of alcohol consumption, of the participant needs to be completed.
- An initial urine sample will be taken and labelled “0 hour” before alcohol and/or NAD dose is consumed.
- Group 1 participants will consume only the vehicle with which the alcohol will be mixed (i.e. lemon flavoured water) over a 15-minute period.
- Group 2 participants will consume a fixed dose of alcohol (Smirnoff vodka 43%, v/v, mixed with lemon flavoured water) of 1.5 g/kg body weight amount over a 15-minute period.
- Group 3 participants will consume the vehicle (i.e. lemon flavoured water), to which 50 mg of NAD has been added, over a 15-minute period.
- Group 4 participants will consume a fixed dose of alcohol (Smirnoff vodka 43%, v/v, mixed with lemon flavoured water) of 1.5 g/kg body weight, to which 50 mg of NAD has been added, over a 15-minute period.
- Urine and saliva for metabolomic investigations will be collected at an agreed time sequence over a period of 5 hours after alcohol and/or NAD dose first consumed, yielding 7 data points (consumption of water 15 minutes before a voiding is needed to facilitate obtaining urine samples).
- The experiment will be done in an environment compatible with clinical and ethical requirements
- A general practitioner will attend the early phases of the experiment (first 2 hours) and will be on call for the remaining part of the experiment.

NOTE: Urine samples must be labelled clearly and precisely and stored in a refrigerated environment and no additional alcohol should be consumed during the experimental time frame.

INFORMED CONSENT PROCEDURES

Participation in the project is fully voluntary. You are free to enquire about the experiment through the investigator and/or supervisor and, if agreeing to participate, the participant will be asked to sign this informed consent form. Should any participant request feedback on the outcomes of this study, such information can be made available to them.

It is required from all participants in this study to complete a questionnaire which provides information which is essential for the project. The participating physician will evaluate the information of all participants, and will approve their participation based on the information given in the questionnaire.

BENEFITS ASSOCIATED WITH THE STUDY

The Master’s study, to which this experiment will contribute, is a research project aimed at better understanding the biological relationship between chronic/acute alcohol consumption and metabolic perturbations within humans. As in all cases, improved knowledge of the normal physiology will eventually yield to a better understanding and treatment of any deviation from normal physiology. The outcome of this research will be used by the researcher for a M.Sc. thesis and no reference will be included in the thesis regarding any individual who participated in the study.

PAYMENT OR REIMBURSEMENT

Participants will not be paid for their participation and do not contribute to the costs of the study. An amount of R100 will however be paid to each of the participants for travel expenses and other inconveniences that resulted from participation in the study as well as R50 for the light meal of choice after completion of the experiment.

CONFIDENTIALITY

All research records are confidential unless the law requires disclosure. No name or other personal identifying information of the participants will be used in any reports or publications resulting from this study. Data from this study will be used in an anonymous statistical analysis and reported as such by the NWU. No patient's identification details will be reported or made known to other parties.

VOLUNTARY PARTICIPATION AND CONDITIONS OF WITHDRAWAL

Your participation in this study is completely voluntary. You may choose not to participate in this study to which you are otherwise entitled.

CONSENT

I, _____, have read and understood the preceding information describing this research study and my questions have been answered to my satisfaction. I voluntarily consent to participate in this research study. I do not waive my legal rights by signing this consent form. I will receive a signed and dated copy of this consent form.

PARTICIPANT:

_____	_____	_____
Printed name	Signature	Date

INVESTIGATOR:

_____	_____	_____
Printed name	Signature	Date

PHYSICIAN:

_____	_____	_____
Printed name	Signature	Date

6. References

1. Cohen, J. *Statistical Power Analysis For Behavioural Sciences*. Second edition. (Hillsdale, NJ: Erlbaum, 1988).
2. Imam, A., Mohammed, U. & Abanyam, C. M. On consistency and limitation of paired t-test, sign and Wilcoxon sign rank test. *IOSR Journal of Mathematics* **10(1)**, 1–6 (2014).
3. Villez, K., Steppe, K. & De Pauw, D. J. W. Use of unfold PCA for on-line plant stress monitoring and sensor failure detection. *Biosystems Engineering* **103**, 23–24 (2009).
4. Irwin, C. *et al.* Contribution towards a metabolite profile of the detoxification of benzoic acid through glycine conjugation: An Intervention Study. *PLOS ONE* **11(12)**, e0167309. doi:10.1371/journal.pone.0167309 (2016).

REFERENCES

- Álvarez-Sánchez, B., Priego-Capote, F. & De Castro, M.D. 2010. Metabolomics analysis I. selection of biological samples and practical aspects preceding sample preparation. *Trends in analytical chemistry*, 29(2):111–119.
- Armstrong, J.A. 2007. Urinalysis in Western culture: a brief history. *Kidney international*, 71(5):384–387.
- Bearn, A.G. & Miller, E.D. 1979. Archibald Garrod and the development of the concept of inborn errors of metabolism. *Bulletin of the history of medicine*, 53(3):315–328.
- Bedair, M. & Sumner, L.W. 2008. Current and emerging mass-spectrometry technologies for metabolomics. *Trends in analytical chemistry*, 27(3):238–250.
- Bérdy, J. 2005. Bioactive microbial metabolites: a personal view. *Journal of antibiotics*, 58(1):1–26.
- Bernini, P., Bertini, I., Luchinat, C., Nincheri, P., Staderini, S. & Turano, P. 2011. Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. *Journal of biomolecular NMR*, 49(3–4):231–243.
- Biology Online's Dictionary. 2005. Artifacts. <http://www.biology-online.org/dictionary/Artifacts>
Date of access: 11 May 2017.
- Blau, N., Duran, M. & Gibson, K.M., eds. 2008. Laboratory guide to the methods in biochemical genetics. Berlin: Springer-Verlag.
- Castle, A.L., Fiehn, O., Kaddurah-Daouk, R. & Lindon, J.C. 2006. Metabolomics standards workshop and the development of international standards for reporting metabolomics experimental results. *Briefings in bioinformatics*, 7(2):159–165.
- Dalgliesh, C.E., Horning, E.C., Horning, M.G., Knox, K.L. & Yarger, K. 1966. A gas-liquid-chromatographic procedure for separating a wide range of metabolites occurring in urine or tissue extracts. *Biochemical journal*, 101(3):792–810.

Dercksen, M., Duran, M., Ijlst, L., Mienie, L.J., Reinecke, C.J., Ruiter, J.P.N., Waterham, H.R. & Wanders, R.J. 2012. Clinical variability of isovaleric acidemia in a genetically homogeneous population. *Journal of inherited metabolic disease*, 35(6):1021–1029.

Dercksen, M., Koekemoer, G., Duran, M., Wanders, R.J.A., Mienie, L.J. & Reinecke, C.J. 2013. Organic acid profile of isovaleric acidemia: a comprehensive metabolomics approach. *Metabolomics*, 9(4):765–777.

Dictionary.com's Dictionary. 2017. Endogenous. <http://www.dictionary.com/browse/endogenous> Date of access: 10 May 2017.

Douglas, F. 2004. GC/MS analysis. *Scientific testimony: an online journal*. <http://www.scientific.org/tutorials/articles/gcms.html> Date of access: 19 Jul. 2016.

Dunn, W.B. & Ellis, D.I. 2005. Metabolomics: current analytical platforms and methodologies. *Trends in analytical chemistry*, 24(4):285–294.

Dunn, W.B., Bailey, N.J.C. & Johnson, H.E. 2005. Measuring the metabolome: current analytical technologies. *Analyst*, 130(5):606–625.

Dunn, W.B., Broadhurst, D., Ellis, D.I., Brown, M., Halsall, A., O'Hagan, S., Spasic, I., Tseng, A. & Kell, D.B. 2008. A GC–TOF–MS study of the stability of serum and urine metabolomes during the UK biobank sample collection and preparation protocols. *International journal of epidemiology*, 37(suppl.1):i23–i30.

Duran, M., Gompertz, D., Bruinvis, L., Ketting, D. & Wadman, S.K. 1978. The variability of metabolite excretion in propionicacidaemia. *Clinica chimica acta*, 82(1–2):93–99.

Duran, M., Dorland, L., Wadman, S.K. & Berger, R. 1994. Group tests for selective screening of inborn errors of metabolism. *European journal of pediatrics*, 153(1):S27–S32.

Eldjarn, L., Jellum, E. & Stokke, O. 1975. Inborn errors of metabolism – a challenge to the clinical chemist. *Clinical chemistry*, 21(1):63–66.

Emwas, A-H.M., Salek, R.M., Griffin, J.L. & Merzaban, J. 2013. NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. *Metabolomics*, 9(5):1048–1072.

Emwas, A-H.M. 2015. The Strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. (In Bjerrum, J.T., ed. *Metabonomics: methods and protocols, methods in molecular biology*. New York, NY: Springer Science & Business Media. v. 1277. p. 161–193).

Erasmus, C., Mienie, L.J., Reinecke, C.J. & Wadman, S.K. 1985. Organic aciduria in late-onset biotin-responsive multiple carboxylase deficiency. *Journal of inherited metabolic disease*, 8:105–106.

Erasmus, E., Mienie, L. J., De Vries, W.N., De Wet, W.J., Carlsson, B. & Larsson, A. 1993. Prenatal analysis in two suspected cases of glutathione synthetase deficiency. *Journal of inherited metabolic disease*, 16(5):837–843.

ERNDIM (European Research Network for evaluation and improvement of screening, Diagnosis and treatment of Inherited disorders of Metabolism). 2017. About us: history of ERNDIM. <http://www.erndim.org/home/about.asp?m=1&s=3> Date of access: 15 Feb. 2017.

Fiehn, O. 2001. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and functional genomics*, 2(3):155–168.

Fiehn, O. 2002. Metabolomics – the link between genotypes and phenotypes. *Plant molecular biology*, 48:155–171.

Fiehn, O., Robertson, D., Griffin, J., Van der Werf, M., Nikolau, B., Morrison, N., Sumner, L.W., Goodacre, R., Hardy, N.W., Taylor, C. & Fostel, J. 2007. The metabolomics standards initiative (MSI). *Metabolomics*, 3(3):175–178.

Fowler, B., Burlina, A. & Vianey-Saban, C. 2008. Quality of analytical performance in inherited metabolic disorders: the role of ERNDIM. *Journal of inherited metabolic disease*, 31(6):680–689.

Free Dictionary's Medical Dictionary. 2003. Biotransformation. <http://www.medical-dictionary.thefreedictionary.com/biotransformation> Date of access: 10 May 2017.

Fukusaki, E. & Kobayashi, A. 2005. Plant metabolomics: potential for practical operation. *Journal of bioscience and bioengineering*, 100(4):347–354.

- Gates, S.C. & Sweeley, C.C. 1978. Quantitative metabolic profiling based on gas chromatography. *Clinical chemistry*, 24(10):1663–1673.
- Gika, H.G., Theodoridis, G.A., Wingate, J.E. & Wilson, I.D.J. 2007. Within-day reproducibility of an HPLC–MS-based method for metabonomic analysis: application to human urine. *Journal of proteome research*, 6(8):3291–3303.
- Gika, H.G., Macpherson, E., Theodoridis, G. & Wilson, I.D.J. 2008. Evaluation of the repeatability of ultra-performance liquid chromatography–TOF–MS for global metabolic profiling of human urine samples. *Journal of chromatography B*, 871(2):299–305.
- Goodacre, R., Broadhurst, D., Smilde, A.K., Kristal, B.S., Baker, J.D., Beger, R., Bessant, C., Connor, S., Capuani, G., Craig, A., Ebbels, T., Kell, D.B., Manetti, C., Newton, J., Paternostro, G., Somorjai, R., Sjöström, M., Trygg, J. & Wlufert, F. 2007. Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics*, 3(3):231–241.
- Griffiths, W.J. & Wang, Y. 2009. Mass spectrometry: from proteomics to metabolomics and lipidomics. *Chemical Society reviews*, 38(7):1882–1896.
- Halket, J.M., Waterman, D., Przyborowska, A.M., Patel, R.K., Fraser, P.D. & Bramley, P.M. 2005. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *Journal of experimental botany*, 56(410):219–243.
- Harrigan, G.G. & Goodacre, R., eds. 2003. Metabolic profiling: its role in biomarker discovery and gene function analysis. New York, NY: Springer Science & Business Media.
- Harris, E.D. 2014. Biochemical facts behind the definition and properties of metabolites. Washington, DC: U.S. Food and Drug Administration. http://www.fda.gov/ohrms/dockets/ac/03/briefing/3942b1_08_harris%20paper.pdf Date of access: 11 May 2017.
- Hendriks, M.M.W.B., Van Eeuwijk, F.A., Jellema, R.H., Westerhuis, J.A., Reijmers, T.H., Hoefsloot, H.C.J. & Smilde, A.K. 2011. Data-processing strategies for metabolomics studies. *Trends in analytical chemistry*, 30(10):1685–1698.
- Holmes E. & Antti, H. 2002. Chemometric contributions to the evolution of metabonomics: mathematical solutions to characterising and interpreting complex biological NMR spectra. *Analyst*, 127(12):1549–1557.

Horning, E.C. & Horning, M.G. 1971. Metabolic profiles: gas-phase methods for analysis of metabolites. *Clinical chemistry*, 17(8):802–809.

Jellum, E. 1977. Profiling of human body fluids in healthy and diseased states using gas chromatography and mass spectrometry, with special reference to organic acids. *Journal of chromatography*, 143(5):427–462.

Jiang, W., Qiu, Y., Ni, Y., Su, M., Jia, W. & Du, X. 2010. An automated data analysis pipeline for GC–TOF–MS metabonomics studies. *Journal of proteome research*, 9(11):5974–5981.

Jonsson, P., Johansson, A.I., Gullberg, J., Trygg, J., Jiye A., Grung, B., Marklund, S., Sjöström, M., Antti, H. & Moritz, T. 2005. High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. *Analytical chemistry*, 77(17):5635–5642.

Kamleh, M.A., Timothy M.D., Ebbels, T.M.D., Spagou, K., Masson, P. & Want, E.J. 2012. Optimizing the use of quality control samples for signal drift correction in large-scale urine metabolic profiling studies. *Analytical chemistry*, 84(6):2670–2677.

Kanani, H., Chrysanthopoulos, P.K. & Klapa, M.I. 2008. Standardizing GC–MS metabolomics. *Journal of chromatography B*, 871(2):191–201.

Kebamo, S., Tesema, S. & Geleta, B. 2015. The role of biotransformation in drug discovery and development. *Journal of drug metabolism and toxicology*, 6(5):196. doi:10.4172/2157-7609.1000196.

Kopka, J. 2006. Current challenges and developments in GC–MS based metabolite profiling technology. *Journal of biotechnology*, 124(1):312–322.

Kouremenos, K.A., Harynuk, J.J., Winniford, W.L., Morrison, P.D. & Marriott, P.J. 2010. One-pot microwave derivatization of target compounds relevant to metabolomics with comprehensive two-dimensional gas chromatography. *Journal of chromatography B*, 878(21):1761–1770.

Lin, L. & Zhang, J. 2017. Role of intestinal microbiota and metabolites on gut homeostasis and human diseases. *BMC immunology*, 18(1):2. PMC5219689. doi:10.1186/s12865-016-0187-3.

Lindon, J.C., Nicholson, J.K., Holmes, E. & Everett, J.R. 2000. Metabonomics: metabolic processes studied by NMR spectroscopy of biofluids. *Concepts in magnetic resonance*, 12(5):289–320.

Little, J.L. 1999. Artifacts in trimethylsilyl derivatization reactions and ways to avoid them. *Journal of chromatography A*, 844(1):1–22.

Lumen Learning in collaboration with Boundless. s.a. Industrial microbiology. <https://courses.lumenlearning.com/boundless-microbiology/chapter/industrial-microbiology/> Date of access: 10 May 2017.

Macherey, A-C. & Dansette, P.M. 2008. Biotransformations leading to toxic metabolites: chemical aspects. (In Wermuth, C.G., ed. *The practice of medicinal chemistry*. 3rd ed. London: Academic Press. p. 674–696).

Mason, S., Moutloatse, G.P., Van Furth, A.M., Solomons, R., Van Reenen, M., Reinecke, C. & Koekemoer, G. 2014. KEMREP: a new qualitative method for the assessment of an analyst's ability to generate a metabolomics data matrix by gas chromatography–mass spectrometry. *Current metabolomics*, 2(1):15–26.

Merriam-Webster Dictionary. 2017. Exogenous. <https://www.merriam-webster.com/dictionary/exogenous> Date of access: 11 May 2017.

Meštrović, T. 2015. What are metabolites? <https://www.news-medical.net/life-sciences/What-are-Metabolites.aspx> Date of access: 10 May 2017.

Nicholson, J.K., Lindon, J.C. & Holmes, E. 1999. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 29(11):1181–1189.

Nicholson, J.K. & Lindon, J.C. 2008. Systems biology: metabonomics. *Nature*, 455(23):1054–1056.

Novotny, M.V., Soini, H.A. & Mechref, Y. 2008. Biochemical individuality reflected in chromatographic, electrophoretic and mass-spectrometric profiles. *Journal of chromatography B*, 866(1):26–47.

Oliver, S.G., Winson, M.K., Kell, D.B. & Baganz, F. 1998. Systematic functional analysis of the yeast genome. *Trends in biotechnology*, 16(9):373–378.

Oxford Living Dictionaries. 2017. Artefact. <https://en.oxforddictionaries.com/definition/artefact>
Date of access: 11 May 2017.

Oxford Reference. 2017. Endogenous substance. <http://www.oxfordreference.com/view/10.1093/oi/authority.20110803095751259#> Date of access: 10 May 2017.

Pasikanti, K.K., Ho, P.C. & Chan, E.C. 2008. Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. *Journal of chromatography B*, 871(2):202–211.

Pauling, L., Robinson, A.B., Teranishi, R. & Cary, P. 1971. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proceedings of the National Academy of Sciences USA*, 68(10):2374–2376.

Peters, V., Garbade, S.F., Langhans, C.D., Hoffmann, G.F., Pollitt, R.J., Downing, M. & Bonham, J.R. 2008. Qualitative urinary organic acid analysis: methodological approaches and performance. *Journal of inherited metabolic disease*, 31(6):690–696.

Preti, G. 2005. Metabolomics comes of age? *The scientist*, 19(11):8.

Quéro, A., Jousse, C., Lequart-Pillon, M., Gontier, E., Guillot, X., Courtois, B., Courtois, J. & Pau-Roblot, C. 2014. Improved stability of TMS derivatives for the robust quantification of plant polar metabolites by gas chromatography–mass spectrometry. *Journal of chromatography B*, 970:36–43.

Reinecke, C.J., Koekemoer, G., Van der Westhuizen, F.H., Louw, R., Lindeque, J.Z., Mienie, L.J. & Smuts, I. 2012. Metabolomics of urinary organic acids in respiratory chain deficiencies in children. *Metabolomics*, 8(2):264–283.

Richards, J.L., Yap, Y.A., McLeod, K.H., Mackay, C.R. & Mariño, E. 2016. Dietary metabolites and the gut microbiota: an alternative approach to control inflammatory and autoimmune diseases. *Clinical & translational immunology*, 5(5):e82. doi:10.1038/cti.2016.29.

Rist, M.J., Muhle-Goll, C., Görling, B., Bub, A., Heissler, S., Watzl, B. & Luy, B. 2013. Influence of freezing and storage procedure on human urine samples in NMR-based metabolomics. *Metabolites*, 3(2):243–258.

Rocca-Serra, P., Salek, R.M., Arita, M., Correa, E., Dayalan, S., Gonzalez-Beltran, A., Ebbels, T., Goodacre, R., Hastings, J., Haug, K., Koulman, A., Nikolski, M., Oresic, M., Sansone, S-A., Schober, D., Smith, J., Steinbeck, C., Viant, M.R. & Neumann, S. 2016. Data standards can boost metabolomics research, and if there is a will, there is a way. *Metabolomics*, 12(1):14–26.

Ryhage, R. 1964. Use of a mass spectrometer as a detector and analyzer for effluents emerging from high temperature gas liquid chromatography columns. *Analytical chemistry*, 36(4):759–764.

Saude, E.J., Adamko, D., Rowe, B.H., Marrie, T. & Sykes, B.D. 2007. Variation of metabolites in normal human urine. *Metabolomics*, 3(4):439–451.

Saudubray, J.M., Sedel, F. & Walter, J.H. 2006. Clinical approach to treatable inborn metabolic diseases: an introduction. *Journal of inherited metabolic disease*, 29(2):261–274.

Scripps Metabolomics Science. s.a. Scripps Metabolomics Science timeline. https://masspec.scripps.edu/landing_page.php?pgcontent=metabolomicScience Date of access: 27 Jan. 2017.

Shurubor, Y.I., Matson, W.R., Willett, W.C., Hankinson, S.E. & Kristal, B.S. 2007. Biological variability dominates and influences analytical variance in HPLC–ECD studies of the human plasma metabolome. *BMC clinical pathology*, 7(9):1. PMC2203971. doi:10.1186/1472-6890-7-9.

Simoni, R.D., Hill, R.L. & Vaughan, M. 2002. Analytical biochemistry: the work of Otto Knuf Olof Folin on blood analysis. *Journal of biological chemistry*, 277(20):19–20.

Smith, P.M., Howitt, M.R., Panikov, N., Michaul, M., Gallini, C.A., Bohlooly-Y, M., Glickman, J.N. & Garrett, W.S. 2013. The microbial metabolites, short-chain fatty acids, regulate colonic T_{reg} cell homeostasis. *Science*, 341(6145):569–573.

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.M., Fiehn, O., Goodacre, R., Griffin, J.L. & Hankemeier, T. 2007. Proposed minimum reporting standards for chemical analysis. *Metabolomics*, 3(3):211–221.

Tanaka, K., Hine, D.G., West-Dull, A. & Lynn, T.B. 1980. Gas-chromatographic method of analysis for urinary organic acids. I. retention indices of 155 metabolically important compounds. *Clinical chemistry*, 26(13):1839–1846.

- Tanaka, K., West-Dull, A., Hine, D.G., Lynn, T.B. & Lowe, T. 1980. Gas-chromatographic method of analysis for urinary organic acids. II. description of the procedure, and its application to diagnosis of patients with organic acidurias. *Clinical chemistry*, 26(13):1847–1853.
- Teahan, O., Gamble, S., Holmes, E., Waxman, J., Nicholson, J.K., Bevan, C. & Keun, H.C. 2006. Impact of analytical bias in metabonomic studies of human blood serum and plasma. *Analytical chemistry*, 78(13):4307–4318.
- Thorburn, A.N., Macia, L & Mackay, C.R. 2014. Diet, metabolites, and “Western-lifestyle” inflammatory diseases. *Immunity*, 40(6):833–842.
- t’Kindt, R., Morreel, K., Deforce, D., Boerjan, W. & Van Bocxlaer, J. 2009. Joint GC–MS and LC–MS platforms for comprehensive plant metabolomics: repeatability and sample pre-treatment. *Journal of chromatography B*, 877(29):3572–3580.
- Van Batenburg, M.F., Coulier, L., Van Eeuwijk, F., Smilde, A.K. & Westerhuis, J.A. 2011. New figures of merit for comprehensive functional genomics data: the metabolomics case. *Analytical chemistry*, 83(9):3267–3274.
- Van den Berg, R.A., Hoefsloot, H.C.J., Westerhuis, J.A., Smilde, A.K. & Van der Werf, M.J. 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.
- Van der Greef, J. & Smilde, A.K. 2005. Symbiosis of chemometrics and metabolomics: past, present, and future. *Journal of chemometrics*, 19(5–7):376–386.
- Van der Greef, J., Van Wietmarschen, H., Van Ommen, B. & Verheij, E. 2013. Looking back into the future: 30 years of metabolomics at TNO. *Mass spectrometry reviews*, 32(5):399–415.
- Villas-Bôas, S.G., Mas, S., Åkesson, M., Smedsgaard, J. & Nielsen, J. 2005. Mass spectrometry in metabolome analysis. *Mass spectrometry reviews*, 24(6):613–646.
- Wadman, S.K., Van der Heiden, C., Ketting, D. & Van Sprang, F.J. 1971. Abnormal tyrosine and phenylalanine metabolism in patients with tyrosyluria and phenylketonuria; gas-liquid chromatographic analysis of urinary metabolites. *Clinica chimica acta*, 34(2):277–287.

Want, E.J., Nordström, A., Morita, H. & Siuzdak, G. 2007. From exogenous to endogenous: the inevitable imprint of mass spectrometry in metabolomics. *Journal of proteome research*, 6(2):459–468.

Want, E.J., Wilson, I.D., Gika, H., Theodoridis, G., Plumb, R.S., Shockcor, J., Holmes, E. & Nicholson, J.K. 2010. Global metabolic profiling procedures for urine using UPLC–MS. *Nature protocols*, 5(6):1005–1018.

Weckwerth, W. & Fiehn, O. 2002. Can we discover novel pathways using metabolomic analysis? *Current opinion in biotechnology*, 13(2):156–160.

Winitz, M., Adams, R.F., Seedman, D.A., Davis, P.N., Jayko, L.G. & Hamilton, J.A. 1970. Studies in metabolic nutrition employing chemically defined diets. 2. effects on gut microflora populations. *American journal of clinical nutrition*, 23:546–559.

Xu, F., Zou, L. & Ong, C.N. 2010. Experiment-originated variations, and multi-peak and multi-origination phenomena in derivatized-based GC–MS metabolomics. *Trends in analytical chemistry*, 29(3):269–280.

Zelena, E., Dunn, W.B., Broadhurst, D., Francis-McIntyre, S., Carroll, K.M., Begley, P., O'Hagan, S., Knowles, J.D., Halsall, A., Wilson, I.D. & Kell, D.B. 2009. Development of a robust and repeatable UPLC–MS method for the long-term metabolomic study of human serum. *Analytical chemistry*, 81(4):1357–1364.

PART IV NEW INSIGHTS ON ACUTE ALCOHOL CONSUMPTION

CHAPTER 6 DISCUSSION

6.1 Reflecting on the aims of my study

6.1.1 The first aim of the study: Provide a holistic view of the comprehensive and complex metabolic consequences that occur due to the consumption of an acute dose of alcohol

The research question that inspired this aim of the study was: “What does the global metabolic profile of acute alcohol consumers (who are not alcoholics) look like, how does it differ from their metabolic profile when no alcohol is consumed, and can metabolomics be used to elucidate these metabolic profiles?”

The second part of the research question, “what do the metabolic profiles of healthy male individuals look like when no alcohol is consumed?”, is an important part of the question, since consumption of the flavoured water is an intervention in itself — its consumption could possibly influence the results from the other interventions and obscure or change the metabolic effects that occur as a result of the alcohol consumption. It was therefore important to determine the effect of consuming the flavoured water. Using NMR spectroscopy, a detailed examination to this end was conducted (Irwin *et al.*, 2016, section 4.4.2). The quantified information of the important metabolites identified in this study, the interpretation of the changes in their excretion over the time-course of the investigation, as well as the biochemical interpretation of the metabolomics data, paved the way for the construction of a metabolite profile reflecting the primary glycine-N-acyltransferase (GLYAT)-catalysed biotransformation reactions of a single xenobiotic (benzoic acid) (Figure 8 of Irwin *et al.*, 2016). The combination of the complementary statistical methods used to resolve the complex NMR spectral data, together with the transdisciplinary approach followed in metabolomics research, provided a more holistic view of the data and the metabolic consequences of ingesting benzoic acid, the preservative in the flavoured water.

From this NMR study, it became evident that hippuric acid, the phase II biotransformation product of benzoic acid, appeared to be the most important discriminating variable in describing the effects of the vehicle-only intervention. In order to determine whether the presence of high concentrations of hippuric acid influences the outcome of the analyses on the alcohol intervention data, the NMR spectral data as well as the quantified GC–MS data were used to

construct lists of important variables. Comparative lists were then constructed after the data pertaining to hippuric acid had been removed from the two data sets. In both the NMR and GC–MS cases, the same relative lists of important metabolites were obtained, leading to the conclusion that hippuric acid does not affect the metabolic effects related to the alcohol intervention (see the Supplementary Information of Irwin *et al.*, 2018 (section 5.7) for details on the GC–MS data comparison).

GC–MS, as well as NMR, were used as analytical platforms to answer the first part of the research question: “What does the global metabolic profile of acute alcohol consumers (who are not alcoholics) look like?”

- The results from the GC–MS investigation (presented in Irwin *et al.*, 2018) revealed significant metabolic effects that occurred due to the consumption of a single, acute dose of alcohol by healthy, moderate-drinking males. Most of these results related to the known alcohol-induced global dysregulation of several metabolic pathways. However, the metabolomics approach implemented in this study offered a novel, convenient and sensitive method to uncover these perturbed metabolic pathways, of which the main ones are illustrated in Figure 5 of Irwin *et al.*, 2018. Using metabolomics, together with various multivariate statistical models, this study provided the first comprehensive description (a holistic view) of the metabolic effects related to acute alcohol consumption.
- The results from the NMR investigation into the metabolic effects of acute alcohol consumption are presented in a second paper, accepted for publication in *PLOS ONE* (section 4.5.2). As typically found in NMR metabolomics studies, the number of quantifiable metabolites is much less than in GC–MS metabolomics investigations (in this case, 14 selected metabolites in the NMR study (Table 2 of the paper accepted for publication in *PLOS ONE*) versus a total data matrix of 120 metabolites in the GC–MS study (Table S2 of the Supplementary Information of Irwin *et al.*, 2018). Nonetheless, a model of the metabolite profile of the important metabolites identified through the NMR analyses (Figure 6 of the paper accepted for publication in *PLOS ONE* gave information on the effect of acute alcohol consumption that was common, as well as complementary, to that in the GC–MS model (Figure 5 of Irwin *et al.*, 2018). This information further contributed to obtaining a holistic view of the metabolic effects related to acute alcohol consumption.

The results and findings presented in the three peer-reviewed papers indicated that metabolomics could be used successfully to elucidate aspects of the metabolic profiles of healthy male individuals both with and without the consumption of alcohol, thereby positively answering the third part of the research question stated above (that is, “can metabolomics be used to elucidate these metabolic profiles?”).

6.1.2 The second aim of the study: Obtain new insights into the biochemical processes and pathways affected by acute alcohol consumption, not observed through the traditional biochemical investigations

Some of the results from both the GC–MS and NMR investigations related to perturbations of metabolic pathways known to be affected by alcohol consumption (reviewed in Chapter 2), including:

- Lactic acid appeared to be a dominant metabolite in the first hour after alcohol consumption in both analyses (for some of the experimental subjects), indicating the early inhibition of aerobic oxidation (Krebs cycle and oxidative phosphorylation) due to the alcohol-induced increased NADH:NAD⁺ ratio. The inhibition of aerobic oxidation is further supported by the observed increase in the excretion of Krebs cycle intermediates (such as fumaric acid, succinic acid, malic acid and 2-hydroxyglutaric acid). This early response was, however, not observed for all the experimental subjects, emphasizing the importance of inter-individual variability in response to alcohol consumption. This variation, which is suggested to be genetically controlled, hints at the apparent genetic predisposition and increased risk of some individuals to become alcohol addicts.
- Increased urinary excretion of hypoxanthine in the early phase following alcohol consumption (observed in the NMR analyses and shown in Table 2 of the paper accepted for publication in *PLOS ONE*) indicates the known ethanol-induced acceleration of adenine nucleotide degradation, and the subsequent increased production of hypoxanthine, xanthine and uric acid.
- Although not measured using either GC–MS or NMR (but mentioned in the paper accepted for publication in *PLOS ONE*), an initial decrease in uric acid excretion was observed following alcohol consumption, possibly due to the inhibition of the NAD⁺-dependent conversion of xanthine to uric acid (a reaction catalysed by xanthine dehydrogenase) by the alcohol-induced increased NADH:NAD⁺ ratio. After approximately 3 hours following alcohol consumption, the levels of uric acid excretion had once again risen, clearly indicating the replenishment of NAD⁺. Interestingly, however, the addition of an NAD supplement (“vehicle plus alcohol and NAD” intervention) did not seem to help alleviate the increased NADH:NAD⁺ ratio, nor the resulting uric acid perturbation. Since NAD is released slowly, taking the supplement only one hour prior to alcohol consumption might not have been sufficient for it to have a noteworthy effect. A follow-up longitudinal study should therefore be performed to determine the correct dosage and timing of NAD supplementation if it is to be used as a potential mechanism of alleviating the effects of acute alcohol consumption. It

is, however, also possible that the observed uric acid profile occurred due to a different mechanism, not at all related to the alcohol-induced NAD^+ depletion.

However, both investigations also gave rise to the identification of important metabolites (such as 2-hydroxyisobutyric acid and sorbitol) that indicated the alcohol-induced perturbation of other metabolic pathways not previously described as being affected by acute alcohol consumption:

- 2-Hydroxyisobutyric acid (2-HIBA), identified only in the GC–MS analysis, is proposed to be increased owing to its reduced reabsorption into the kidneys. This is a result of the increased competition with alcohol-induced increased lactic acid for reabsorption through a common renal monocarboxylic acid transporter (described in Irwin *et al.*, 2018). In addition, 2-HIBA has been shown to be an acylation agent of histones, opening the question of a possible epigenetic consequence of acute alcohol consumption.
- Sorbitol is proposed to be increased either due to its being one of several osmolytes (another one is trimethylamine-N-oxide (TMAO)) that exit hepatocytes (and are subsequently excreted in the urine) in response to the apparent influx of ethanol, or due to its conversion to fructose (in an NAD^+ -dependent reaction catalysed by sorbitol dehydrogenase) being inhibited by the alcohol-induced increased $\text{NADH}:\text{NAD}^+$ ratio (see Figure 6 of the paper accepted for publication in *PLOS ONE*). The increased urinary excretion of sorbitol, whether due to osmoregulation or in response to the altered redox state in the hepatocytes, was observed only in the NMR analysis.

To summarize, Figure 6–1, an amended version of Figure 2–2 — which illustrates the global perturbations related to the alcohol-induced increased $\text{NADH}:\text{NAD}^+$ ratio — illustrates the known effects observed during my study, as well as these new insights into the consequences of acute alcohol consumption.

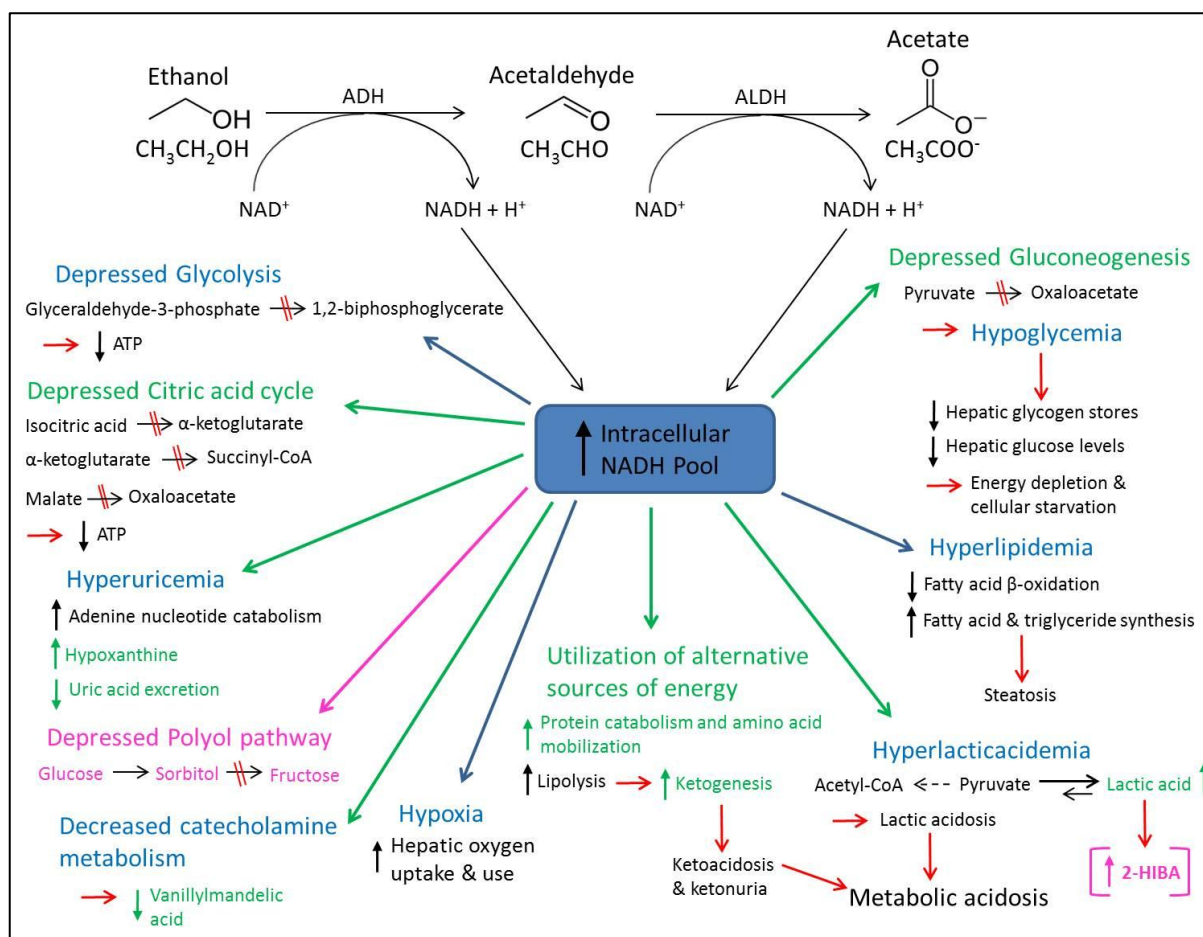


Figure 6-1: Revised and expanded version of the model of the global metabolic perturbations induced by the alcohol-induced increased NADH:NAD^+ ratio (presented as Figure 2–2 in Chapter 2). The known effects observed, and the new insights gained, through my study are indicated in green and pink, respectively.

It is important to note that some commonality, but also some complementarity, was revealed by the combination of the untargeted (NMR) and semi-targeted (GC–MS) approaches. Both approaches indicated important consequences of the hepatic ADH- and ALDH-catalysed catabolism of consumed ethanol, such as the notably high early excretion of lactic acid. In addition, the role of GLYAT-catalysed biotransformation of benzoic acid also appeared as a dominant observation in both approaches. Complementary results included the observation of Krebs cycle dysfunction and branched-chain amino acid catabolism (strongly supported only by the GC–MS data), as well as purine catabolism and osmoregulation through sorbitol efflux (strongly supported only by the NMR data). Also, it is important to note that neither of these analytical approaches indicated perturbations in the lipid metabolism, which is known to be affected by alcohol consumption. In order to observe these perturbations, however, analysis using an LC-based platform (lipidomics) would be the approach of choice.

6.2 Lessons learned through this study

The experimental design for my study involved the collection of a very large number of urine samples, summarized in Figure 6–2. The addition of a large number of QC samples, as well as repeat samples, for the evaluation of repeatability and to account for the various sources of variation that is known to exist in GC–MS analyses (Chapter 5), resulted in the need to prepare and analyse more than 1000 samples. As described in Chapter 5, all 24 batches of samples were analysed in a continuous fashion, and the results from each batch were stored electronically in the GC–MS database until all the analyses from all 24 batches were completed. Only then were all the results from all 24 batches evaluated, analysed and interpreted simultaneously.

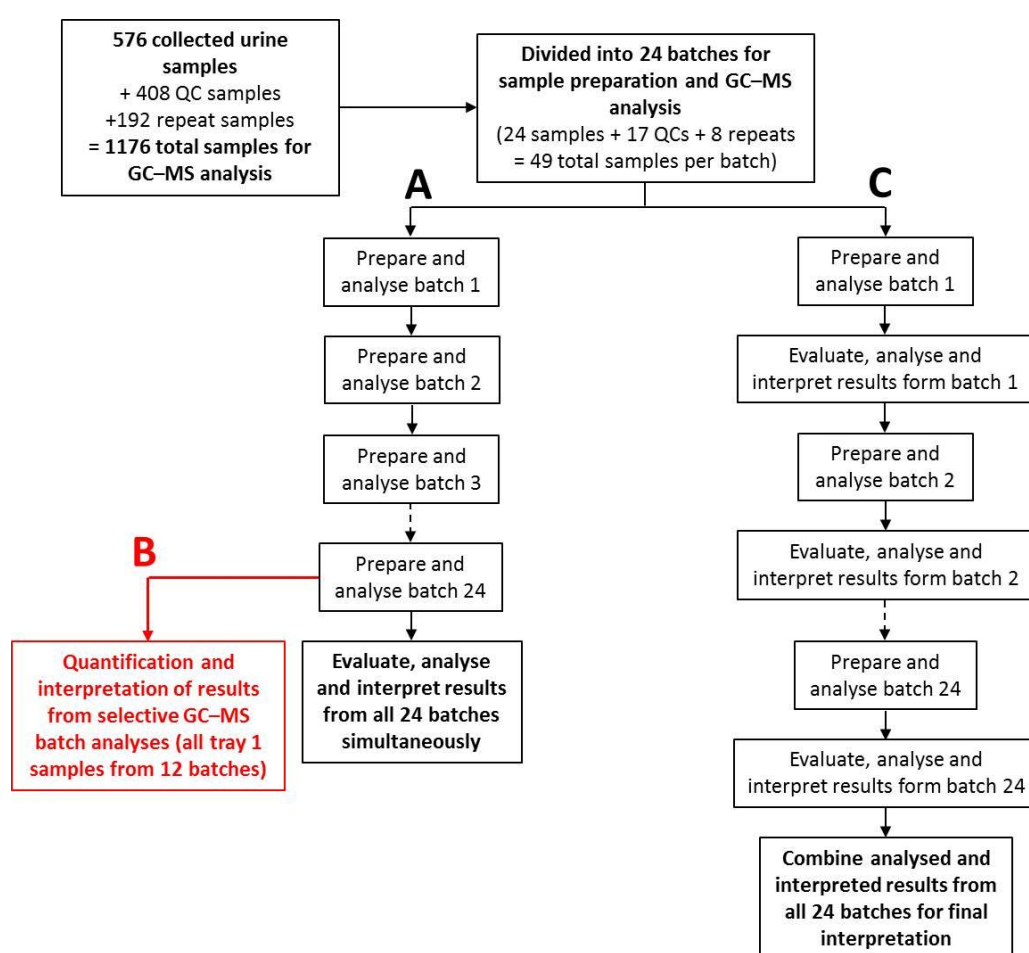


Figure 6-2: Illustration of two main methods of undertaking a study such as this one with regard to GC–MS sample preparation and analysis, and interpretation of the results. (A) The method intended to be implemented in my study; (B) the method, an amendment to method A, actually implemented in my study, consisting of analysing data obtained from the GC–MS analyses of samples from selected cases only; (C) the proposed alternative, preferred method.

The advantage of analysing samples in this way (Figure 6–2 A) is that it is quick to analyse such a large number of samples and get the experimental analyses out of the way so that focus can be placed on analysis and interpretation of the results. However, this method has an extreme disadvantage, which became evident during my study and resulted in the decision to not use a large part of the data — results are only evaluated, analysed and interpreted after completion of all experimental analyses, which means that no problems or inconsistencies in the data can be picked up or known before this. Consequently, no adjustments possibly to improve the results of inconsistent batches can be made to the method of sample analysis unless the entire analysis was redone.

Using this method had some detrimental consequences, including unrepeatable results and many quantification inconsistencies. It was also from these data that the inconsistencies in the data obtained from the analyses of the samples in the second tray of each batch became evident, leading to the eventual decision to not use these data (section 5.6.2). This was not ideal, since it meant having no GC–MS data pertaining to the NAD interventions, but was a necessary sacrifice because, due to time constraints, the samples could not all be reanalysed in an attempt possibly to obtain better, more repeatable results. For these reasons, this method of analysis could not be fully implemented and an alternative last step needed to be found.

Fortunately, the data from the samples from the first tray of most of the batches showed less variation in terms of repeatability (illustrated in Figure 5–5), and could therefore still be used. Thus, the method illustrated in Figure 6–2 A was slightly amended to produce the approach illustrated in Figure 6–2 B — instead of evaluating, analysing and interpreting the results from all 24 batches simultaneously, variable identification and quantification were done only for all tray 1 samples (that is, the “vehicle only” and “alcohol plus vehicle” interventions) from 12 of the batches. The same method that was used to identify the 172 variables in the QC samples (detailed in section 5.5), was used to identify and quantify these same variables in these selected experimental samples. This method of meticulously quantifying the metabolites in each sample individually is highly reliable, but also extremely tedious and time-consuming, and therefore only samples from 12 of the 24 batches were analysed in this way. The results obtained from identifying and quantifying metabolites in this manner, however, seemed promising, and could be used for the metabolomics analysis of the effect of alcohol consumption (presented in Irwin *et al.*, 2018). The advantage of this method of identifying and quantifying variables is that additional variables, formed in the side-reactions that typically occur during chemical derivatization, are observed. Also, through the use of this method, the analyst acquires refined technical skills regarding deconvolution, improved ability on spectral selection from the available MS libraries, and insight into the relationship between structure (MS data) and retention time (GC data). All in all, this method is highly recommended for analyses of “the

single case”, such as analyses to test for inborn errors of metabolism, or in small-scale metabolomics studies (cases and control groups of approximately 20 cases). Owing to the time-consuming nature of data analysis, however, this method is not recommended for large-scale metabolomics investigations.

From my experience, I would propose another, possibly better, way of doing things — evaluating, analysing and interpreting the results from each batch of samples separately and directly following GC–MS analysis of that batch, before sample preparation and analysis of the next batch is started. I regard this proposed alternative method (Figure 6–2 C) to be the preferred choice, since it holds major advantages over the method illustrated in Figure 6–2 A by allowing for many more amendments and optimisation in subsequent batches to obtain the best results. If, upon analysing and interpreting results from a particular batch, discrepancies in repeatability are encountered, batch corrections can be performed, or the analysis of that batch can be repeated in order to try to improve results. If it is observed that repeatability is a constant issue across multiple batches — or, as in this case, towards the end of each batch — it may be that the batches are too long and need restructuring. The decision can then be made, for example, to analyse samples one tray at a time, instead of one batch at a time, which, one hopes, would give more repeatable results. Another advantage of this alternative method is the direct involvement of the analyst in the assessment of the quality of the data generated from each batch, thereby providing the opportunity to improve his/her analytical skills.

Another alternative method, intermediate between approaches A and C in Figure 6–2, that may be considered when working with analytically reasonable GC–MS data, is to identify some key assumptions in the batch just analysed. These could include comparing the internal standard levels of all samples and checking the data from all QC samples in each batch after they have been analysed. These preliminary assessments would provide a basic idea of the quality of the data from that batch. Analysis of data from other variable peaks should, however, still be done only when data from all batches are available. This approach requires advanced computer skills to perform the data analysis and corrections, but would enhance the analyst’s insights into the character and ways of improving metabolomics data.

Being able to account for problems and to adjust the method of sample analysis during the analysis avoids the possible need to repeat the entire analysis if results are found to be without value. This would save time and reagents, and would result in obtaining the best possible data set. However, the amount of trial and error, and repeat analyses of samples and batches that may be involved in optimizing the method proposed in Figure 6–2 C for it to be implemented efficiently, would (1) be extremely time-consuming, (2) require additional research funding, as well as (3) laboratory equipment (GC–MS apparatus), preferably dedicated to a single mode of analysis, and (4) involve dedicated analysts. These are luxuries that I did not have at my

disposal, and are unlikely to be affordable for research in a developmental setting such as the one we experience in South Africa. These realities obliged me to use the first method.

However, by no means does this imply that metabolomics research in South Africa will always be mediocre, incremental and below par due to the developmental status of the country. The lessons learned from my experience first bring meticulous *planning* of an important investigation into focus. The experimental design of this study, and the associated measurement designs, were aimed at addressing too many interventional and theoretical issues in one experiment — (1) biological effects of four different experimental interventions; (2) analytical information on repeatability through the addition of repeat samples, and the value of randomized sample distribution per tray; and (3) assessment of repeatability through the inclusion of QC samples, providing information for statistical methods of batch corrections. Additionally, overriding these empirical aspects, the ambition to develop a statistical model and simulated data for the experimental design, which would provide an opportunity for comparison of real and theoretical metabolomics information, was also proposed. This approach was over-ambitious and neglected the wisdom of simplification, a concept beautifully articulated by the theologian Cornelius Plantinga Jr. when he said: “simplicities lie beyond complexities and incorporate them” (Plantinga, 1998). Moreover, much is made of the multidisciplinary approach involving biologists, chemists and statisticians that is required for metabolomics investigations. From my experience, I suggest that the analyst should be included as an essential participant from the initial planning to the final implementation stages of a metabolomics investigation.

Second, the unfortunate and partly detrimental results of my study have placed the need for revision and improvement of *standard laboratory practice* into focus. Following my study, many changes have been incorporated into the SOPs and sample handling and analytical procedures at PLIEM. For example, (1) sample batches for GC–MS analysis should preferably not exceed 10 samples, and may not be so many that the analyses run for more than 24 hours; and (2) derivatizing samples at a lower temperature for a longer period (45 minutes at 75°C instead of 30 minutes at 80°C) seems to improve repeatability and prevent excessive artifact formation. Additionally, specific procedures and regulations that need to be implemented and adhered to regarding the moving of laboratory equipment have been put in place. One hopes that these changes can prevent problems and situations similar to the ones encountered during my study from recurring.

Third, possibly because it is a developing country, South Africa provides a wealth of societal and environmental issues and realities for novel investigations using metabolomics technology. These investigations could, however, benefit immensely from increased *co-operation* from governmental agencies — already evident from our contractual co-operation with the Technological Innovation Agency (TIA) of the South African Department of Science and

Technology — as well as scientific communities from different countries. Prime examples of such collaborative approaches can be provided by our own laboratory, illustrated by (1) the participation of PLIEM in the ERNDIM quality assurance schemes (Chapter 5); (2) the allocation of SKILLS stipends, an initiative introduced by the Dutch Embassy in Pretoria, to some of our Ph.D. students to enable them to collaborate with researchers in The Netherlands; (3) co-authorship of most of our metabolomics publications with academics in The Netherlands; and finally (4) the appointment of my co-promoter as Professor Extraordinary at NWU.

6.3 Future prospects

The findings from my study opened various avenues for potentially important future investigations in alcohol research, of which a few are mentioned in the following sections.

6.3.1 2-Hydroxyisobutyric acid

The striking presence of 2-HIBA observed in the GC–MS analyses, detailed in Irwin *et al.*, 2018, supports the emerging new paradigm of 2-HIBA being an important, now considered an endogenous, metabolite. This compound, which has been related to various disease conditions, was also observed to be the most important metabolite that discriminates patients with Fibromyalgia Syndrome from controls in a Ph.D. study by a fellow student in our laboratory (Malatji, 2017), further emphasizing its potential importance and relation to diseases. Future studies to determine its biological origin, as well as its involvement in alcohol consumption and its perceived gene-modification role through lysine 2-hydroxyisobutyrylation of histones may take us one step closer to understanding the personalized responses to acute alcohol consumption and the perceived epigenetic changes that it induces. This could help researchers to possibly define novel approaches to treat or ameliorate alcohol-induced disability, organ damage and morbidity.

6.3.2 Biomarker identification

Biomarkers are a small subset of metabolites that are objectively measured and evaluated as indicators of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention (Biomarkers Definitions Working Group, 2001; Hendriks *et al.*, 2011).

Currently, there are no absolutely reliable biochemical measures of alcohol consumption, which is a big obstacle to the diagnosis and treatment of, as well as research on, alcoholism. Testing for ethanol itself in urine, breath or blood is inadequate, since alcohol is rapidly metabolized within hours after use, making it an insensitive marker to establish abstinence unequivocally (Wurst *et al.*, 2005). Various alternative metabolites (including direct ethanol metabolites and

metabolites that are measures of alcohol-induced tissue damage), with varying degrees of sensitivity, specificity and reliability, have been proposed as promising biomarkers for alcohol consumption (Freeman & Vrana, 2010; Litten *et al.*, 2010; Wurst *et al.*, 2005). A prime candidate is ethyl glucuronide, a long-term marker of alcohol consumption, which can be measured in blood (for 14–24 hours), in urine (for 3–5 days), and in hair (for several months) after alcohol consumption has ceased. It is very sensitive and specific to alcohol consumption, and is an important marker evaluated by liver transplant committees (Litten *et al.*, 2010; Skipper *et al.*, 2004; Wurst *et al.*, 2005).

The successful development of specific, sensitive, reliable, simple and inexpensive biomarkers that allow for the accurate assessment of alcohol consumption and drinking patterns would be valuable for more accurate screening and monitoring of drinking, for early detection and prevention of alcoholism and its associated diseases, and could significantly improve therapy effectiveness and treatment outcome, leading to health, social and socio-economic benefits. Additionally, these biomarkers could be used to measure drinking outcome during clinical trials, as well as to identify recent drinking in high-risk situations (Freeman & Vrana, 2010; Litten, *et al.*, 2010; Manna *et al.*, 2010; Wurst *et al.*, 2005).

Metabolomics can play a crucial role in the development of such biomarkers — by analysing the metabolites arising from perturbed metabolic pathways, metabolomics approaches can be used to understand the mechanisms that cause alcohol-induced metabolic perturbations and tissue injury, and to define the roles of the different metabolic pathways affected by alcohol abuse. This understanding could contribute to the identification of biomarkers of alcohol use and abuse, and to the development of novel medications for the treatment of alcoholism and alcohol-related conditions (Gunzerath *et al.*, 2011; Harrigan *et al.*, 2008; Wurst *et al.*, 2005).

During the GC–MS analysis, vanillylmandelic acid emerged as a candidate biomarker for alcohol consumption in seemingly non-intoxicated individuals following the consumption of an acute dose of alcohol (detailed in Irwin *et al.*, 2018). However, the full development and validation of new biomarkers is not a simple process, and is far beyond the scope of my study, but further research into their development, with the use of metabolomics, should definitely be pursued in the future.

A third, truly interesting and important factor to consider for any future investigations into the effects of alcohol consumption, is the consideration of the various genetic polymorphisms of the genes responsible for encoding the alcohol-metabolizing enzymes ADH and ALDH. This future prospect is discussed in more detail in the following section.

6.3.3 Genetic polymorphisms

First, a brief overview of the link between polymorphisms and alcoholism is given in order to understand the importance of this factor for future alcohol-related research.

6.3.3.1 What is a polymorphism?

“Genetic polymorphisms”, a term first defined in 1940 by E.B. Ford (Ford, 1940), are heritable differences in a DNA sequence, at the same place on a chromosome, between members of the same biological species. When these differences, or genetic variations, occur in enzyme-coding sections of DNA, they result in the production of enzymes with slight variations (called isoforms). These differences can result in isoforms having varying degrees of enzymatic or catalytic activity in metabolising systems (Hopkinson & Whitehouse, 2000; Satoh, 2007), a characteristic that frequently underlies the differences in individual susceptibility to, or protection from, diseases and toxicants such as alcohol (Ferrer-Dufol & Menao-Guillen, 2009; Gunzerath *et al.*, 2011). The severity of illness and response to treatments are also manifestations of such genetic variations (Manna *et al.*, 2011).

6.3.3.2 The effect of polymorphisms on alcohol consumption

Alcohol drinking habits, the response to alcohol consumption, and the vulnerability to becoming alcohol-dependent and to developing alcohol-related organ damage varies greatly between different populations, and even among individuals within the same population (Manna *et al.*, 2011). Also, it is known, and was observed in my study, that large variations in the rate of alcohol metabolism and alcohol elimination efficiency are manifest among individuals of a homogeneous population, when equivalent amounts of alcohol are consumed (Agarwal, 2001; Zakhari & Li, 2007). Genetic background and composition, including polymorphisms, are determining factors for explaining these variations (Ferrer-Dufol & Menao-Guillen, 2009). The different genetic polymorphisms of the genes encoding the alcohol-metabolising enzymes ADH and ALDH, which exist between populations and individuals, result in significant differences in the activities of these enzymes, and variations in their abilities to metabolize alcohol and acetaldehyde (Chen, C-C. *et al.*, 1999; Chen *et al.*, 2009; Lieber, 1997). According to the literature, only the genes encoding these enzymes have been firmly established to influence an individual's vulnerability to alcoholism and the possible risk for developing ALD (Chen, C-C. *et al.*, 1999; Zakhari & Li, 2007).

ADH polymorphisms

ADH enzymes metabolize a wide variety of substrates, including ethanol and other toxic agents, and are encoded by a polygenic family of seven genes. These genes are divided into five classes according to the subunit and physicochemical and kinetic characteristics of the

enzymes they encode (Lieber, 1997). Over 20 ADH isoforms, which vary in their amino acid composition, primary sequence, substrate specificity and affinity, maximal oxidation rate, and organ distribution, are known (Agarwal, 2001; Mardh *et al.*, 1985). However, to date, only the polymorphisms of the class I ADH2 and ADH3 genes seem to noticeably affect alcohol metabolism and dependence. The ADH2 gene is responsible for most of the ADH activity on ethanol in the liver, and has three polymorphisms, namely ADH2*1, ADH 2*2 and ADH2*3. The ADH3 gene has two polymorphisms, namely, ADH3*1 and ADH3*2 (Agarwal, 2001; Chen *et al.*, 2009; Ferrer-Dufol & Menao-Guillen, 2009; Lieber, 1997; Zakhari & Li, 2007).

ALDH polymorphisms

The human ALDH superfamily encodes more than 30 ALDH isoforms that metabolize a wide spectrum of endogenous and exogenous aldehydes, and differ in their electrophoretic mobility, kinetic properties, and cellular and tissue distribution (Agarwal, 2001; Chen *et al.*, 2009). Cytosolic ALDH1 and mitochondrial ALDH2 primarily contribute to the oxidation of acetaldehyde (Chen, C-C. *et al.*, 1999; Chen *et al.*, 2009; Ferrer-Dufol & Menao-Guillen, 2009), but only two of the four polymorphisms of the ALDH2 gene, namely, ALDH2*1 and ALDH2*2, seem to noticeably affect alcohol metabolism (Agarwal, 2001). Under normal physiological conditions, ALDH3 does not metabolize acetaldehyde efficiently (Chang *et al.*, 2017), and seems to play no significant role in ethanol metabolism (Ferrer-Dufol & Menao-Guillen, 2009). Therefore, the polymorphisms of ALDH3 (ALDH3*1 and ALDH3*2) should have little or no effect on alcohol metabolism (Agarwal, 2001).

Many studies have suggested that the different polymorphisms of the ADH2, ADH3 and ALDH2 genes are associated with the differences in an individual's risk of developing alcoholism and ALD (Chen, C-C. *et al.*, 1999; Chen, Y-C. *et al.*, 1999; Gunzerath *et al.*, 2011; Zintzaras *et al.*, 2006). The definitive characterisation of putative alcoholism vulnerability genes and polymorphisms may be useful in improving preventative and treatment approaches for alcoholism (Agarwal, 2001). There are a total of 12 different combinations of the polymorphisms of these genes. Depending on the combination, individuals are more or less susceptible to the development of alcoholism, since each of these polymorphisms either protects individuals from the development of alcoholism or makes them more susceptible to it (Agarwal, 2001; Chen, C-C. *et al.*, 1999; Chen, Y-C. *et al.*, 1999; Chen *et al.*, 2009; Ferrer-Dufol & Menao-Guillen, 2009; Gunzerath *et al.*, 2011; Peng *et al.*, 1999; Zakhari & Li, 2007). It is suggested that people with a gene combination of ADH2*2 or ADH2*3, ADH3*1 and ALDH2*2 are genetically the most protected from the development of alcoholism and alcohol-related diseases (Agarwal, 2001; Chen, C-C. *et al.*, 1999; Gunzerath *et al.*, 2011), whereas people with a gene combination of ADH2*1, ADH3*2 and ALDH2*1 are genetically the most susceptible (Zintzaras *et al.*, 2006).

Table 6–1 shows the ADH and ALDH genes related to alcohol metabolism, as well as their various polymorphisms and their effects on the development of alcoholism.

Table 6-1: Details of the various ADH and ALDH genes related to alcohol metabolism

Enzyme	Class	Gene	Location of enzymatic action (affinity for ethanol)	Polymorphism	Nucleotide position	Amino acid change	Population distribution	Effect on development of alcoholism
ADH	I	ADH1	Hepatic (High)			Wild type		
		ADH2	Hepatic (High)	ADH2*1		Wild type	> 90% Caucasians	Susceptible
				ADH2*2	47	Arg > His	85% Asians 5–10% Caucasians Virtually absent in Africans	Protective
				ADH2*3	369	Arg > Cys	15–20% Africans Rare in Caucasians and Asians	Protective
		ADH3	Hepatic (High) Gastric	ADH3*1		Wild type	> 90% Africans and Asians 60% Caucasians	Protective
				ADH3*2	271 & 349	Arg > Gln & Ile > Val	Very rare in Africans and Asians 40% Caucasians	Susceptible
	II	ADH4	Hepatic (Low)			Wild type		
	III	ADH5	Gastrointestinal (first-pass) (Low)			Wild type		
	IV	ADH7	Gastric			Wild type		
	V	ADH6	Physiological functions and kinetic properties unknown			Gene product unknown		
ALDH	I	ALDH1	Hepatic (cytosol)			Wild type	Rare but have been described	Unknown
	II	ALDH2	Hepatic (mitochondria)	ALDH2*1		Wild type	Most Caucasians < 50% Africans and Asians	Susceptible
				ALDH2*2	487	Glu > Lys	50% Asians Virtually absent in Caucasians and Africans	Protective
				ALDH2*3	479	Glu > Lys		Unknown
				ALDH2*4		Silent		Unknown
		ALDH3	Gastric	ALDH3*1		Wild type		None
				ALDH3*2	329	Pro > Ala		None

An understanding of the role of the quantity and frequency of alcohol consumed, as well as that of the genetic variations related to alcohol metabolism, is essential for appreciating both the short-term and long-term effects of alcohol in liver pathogenesis, and to optimize the approaches to patient treatment (Zakhari & Li, 2007). Also, studies suggest that the relatives of alcoholics have a greater risk of developing alcohol dependence (Gunzerath *et al.*, 2011). It is therefore important to study the genetic factors that increase a person's risk of developing alcoholism, so that precautionary measures and/or medication could possibly be developed to prevent susceptible individuals from developing the condition.

6.3.3.3 Future studies on acute alcohol consumption

From this overview it is evident that the different polymorphisms of the genes encoding the alcohol-metabolizing enzymes significantly affect the rate and efficiency of alcohol metabolism. Many studies to determine the correlation between the susceptibility to develop alcoholism and these polymorphisms have been performed. However, no such study on acute alcohol consumption has been reported. A study of this magnitude is far beyond the scope of my project, but, based on the apparent importance of these polymorphisms, I believe that it is important and would be advantageous, in the future, to conduct a metabolomics study similar to the one presented in this thesis, which takes the aspect of the different polymorphisms into account. Below is a proposed experimental design for such a future investigation:

First of all, it is of critical importance to establish whether any of these different polymorphisms — and then which ones — would play a role in ethanol and acetaldehyde metabolism following acute alcohol consumption. Based on the large inter-individual variations observed in my study, I suspect that they would, but it would be of great value to consult with experts in the field of alcohol, alcohol metabolism and the related polymorphisms, in order to confirm this.

If these suspicions are confirmed, the first experimental step in this proposed investigation would involve genotyping — the characterization of a person's genotype (the exact description of the complete set of genes of an individual) with respect to a particular genetic variant (Smith, 2002) — of the ADH and ALDH genes of all experimental subjects, in order to determine the exact ADH and ALDH isoforms of each subject. In order to exclude additional confounding factors, all experimental subjects to be tested should be selected from a population that is as homogeneous as possible. Subjects with the same genotype would then be grouped together to create study groups that are phenotypically, as well as genotypically, as homogeneous as possible. However, owing to the large number of known polymorphisms, and the extreme rarity of some of these polymorphisms (especially in certain populations), an extremely large cohort of study subjects would be required for

these initial genotype screenings, to obtain a sufficient number of subjects with each of the different polymorphisms, so that each one is sufficiently represented. Ideally, between 12 and 24 experimental subjects per group would probably be required, but many (possibly hundreds or thousands) more would need to be screened. Procuring such a large number of experimental volunteers is a potentially crippling obstacle for a study such as the one proposed here.

The next aspect to consider is the choice of analytical platform. The application of GC–MS and NMR in my study produced informative results of high quality. However, it might be that other platforms, such as LC–MS for example, could produce even better results and reveal metabolites that are not able to be measured by either GC–MS or NMR. The identification and quantification of ethyl sulphate, a recently confirmed human metabolite of ethanol, is one example of a rare metabolite that should be included in a metabolism study involving the polymorphisms of alcohol-metabolizing enzymes (Mitchell *et al.*, 2014). Once again, consulting experts would be highly beneficial in this respect. Despite the choice of analytical platform, I would still suggest using a metabolomics approach, both untargeted and targeted, such as I employed in my study. Such an approach would be advantageous, since these analyses can help to capture a wide-angle view of the varying biochemical responses to alcohol consumption that exist between individuals and populations as a result of the variability in the ability of the different ADH and ALDH isoforms to metabolize ethanol and acetaldehyde (Manna *et al.*, 2011), as well as of the individual characteristics of the time-dependence of ethanol metabolism.

The rest of the study would follow much the same line as my study — by involving a cross-over design in which all experimental subjects from each of the different polymorphic groups take part in each of the interventions. In this way, each subject can act as its own control, thereby making it much easier to attribute the results obtained to a certain intervention. For this future study, the interventions would probably consist of (1) the consumption of a vehicle, and (2) the consumption of an acute dose of alcohol, together with the same vehicle. This would permit the possible differences in the rates and efficiencies of ethanol and acetaldehyde metabolism — due to the different polymorphisms in the different polymorphic groups — to be easily observable.

The large number of individuals who would have to be tested in order to obtain sufficient numbers of experimental subjects would also add various ethical considerations. Therefore, obtaining ethical approval for a study such as the one proposed here would be extremely difficult, or maybe even impossible, and is another large obstacle that would need to be overcome.

Envisaging a study of this magnitude, and involving this many factors and ethical constraints, is probably quite idealistic, and would in all likelihood never be feasible to carry out. However, if at all possible, I think such a study would add great value to further the understanding of the effects of alcohol consumption, and could possibly even contribute to the identification of biomarkers of alcohol consumption or the development of effective treatments to prevent and overcome alcohol abuse and alcohol dependence.

Another point of interest would be to perform an in depth investigation of the detoxification pathways following acute alcohol consumption, with specific focus on the acylcarnitines. Ideally, serum samples would be used for such an investigation, and samples would be analysed using LC–MS/MS, which is currently the optimal method for acylcarnitine analysis. Herein lies an auspicious, potential follow-up study (possibly a post-doctoral study), incorporating a third analytical metabolomics platform in order to profile acylcarnitines following acute alcohol consumption.

One final note: a recent publication on alcohol use disorder (AUD) emphasized personality as an intermediate phenotype in the genetic dissection towards alcohol use and the development of alcoholism. The authors' final conclusion is a reminder of the importance of reflection on the humane dimensions of dealing with conditions as serious as alcohol use and abuse (Oreland *et al.*, 2017):

“AUD can be a consequence of an impulsive, sensation-seeking and extroverted personality, as well as of an anxiety-prone and introverted personality, which by several neurobiological markers might represent two different extremes with regard to personality. This complexity of AUD is reflected in many efforts through the years to construct psychological models to classify clinically relevant AUD subtypes to improve the clinical management. Personality is strongly regulated by genetic factors and should therefore be a top candidate intermediate correlate for the dissection of the genetic underpinnings of different subtypes of AUD.”

REFERENCES

- Agarwal, D.P. 2001. Genetic polymorphisms of alcohol metabolizing enzymes. *Pathologie biologie*, 49(9):703–709.
- Biomarkers Definitions Working Group. 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical pharmacology & therapeutics*, 69:89–95.
- Chang, J.S., Hsiao, J-R. & Chen, C-H. 2017. ALDH2 polymorphism and alcohol-related cancers in Asians: a public health perspective. *Journal of biomedical science*, 24(1):19. doi:10.1186/s12929-017-0327-y.
- Chen, C-C., Lu, R-B., Chen, Y-C., Wang, M-F., Chang, Y-C., Li, T-K. & Yin S-J. 1999. Interaction between the functional polymorphisms of the alcohol-metabolism genes in protection against alcoholism. *American journal of human genetics*, 65(3):795–807.
- Chen, Y-C., Lu, R-B., Peng, G-S., Wang, M-F., Wang, H-K., Ko, H-C., Chang, Y-C., Lu, J-J., Li, T-K. & Yin, S-J. 1999. Alcohol metabolism and cardiovascular response in an alcoholic patient homozygous for the ALDH2*2 variant gene allele. *Alcoholism: clinical and experimental research*, 23(12):1853–1860.
- Chen, Y-C., Peng, G-S., Wang, M-F., Tsao, T-P. & Yin, S-J. 2009. Polymorphism of ethanol-metabolism genes and alcoholism: correlation of allelic variations with the pharmacokinetic and pharmacodynamic consequences. *Chemico-biological interactions*, 178(2):2–7.
- Ferrer-Dufol, A. & Menao-Guillen, S. 2009. Toxicogenomics and clinical toxicology: an example of the connection between basic and applied sciences. *Toxicology letters*, 186(1):2–8.
- Ford, E.B. 1940. Polymorphism and taxonomy. (In Huxley, J., ed. *The new systematics*. Oxford: Clarendon Press. p. 493–513).
- Freeman, W.M. & Vrana, K.E. 2010. Future prospects for biomarkers of alcohol consumption and alcohol-induced disorders. *Alcoholism: clinical and experimental research*, 34(6):946–954.
- Gunzerath, L., Hewitt, B.G., Li, T-K. & Warren, K.R. 2011. Alcohol research: past, present, and future. *Annals of the New York Academy of Sciences*, 1216(1):1–23.

Harrigan, G.G., Maguire, G. & Boros, L. 2008. Metabolomics in alcohol research and drug development. *Alcohol research & health*, 31(1):26–35.

Hendriks, M.M.W.B., Van Eeuwijk, F.A., Jellema, R.H., Westerhuis, J.A., Reijmers, T.H., Hoefsloot, H.C.J. & Smilde, A.K. 2011. Data-processing strategies for metabolomics studies. *Trends in analytical chemistry*, 30(10):1685–1698.

Hopkinson, D. & Whitehouse, D. 2000. An introduction to genetic polymorphism. (In King, M-J., ed. Human blood cells: consequences of genetic polymorphisms and variabilities. London: Imperial College Press. p. 1–14).

Lieber, C.S. 1997. Ethanol metabolism, cirrhosis and alcoholism. *Clinica chimica acta*, 257(1):59–84.

Litten, R.Z., Bradley, A.M. & Moss, H.B. 2010. Alcohol biomarkers in applied settings: recent advances and future research opportunities. *Alcoholism: clinical and experimental research*, 34(6):955–967.

Malatji, B. 2017. A metabolomics investigation of fibromyalgia syndrome. Potchefstroom: NWU. (Thesis – Ph.D.). (Submitted for examination).

Manna, S.K., Patterson, A.D., Yang, Q., Krausz, K.W., Li, H., Ilde, J.R., Fornace Jr., A.J. & Gonzalez, F.J. 2010. Identification of noninvasive biomarkers for alcohol-induced liver disease using urinary metabolomics and the *Ppara*-null mouse. *Journal of proteome research*, 9(8):4176–4188.

Manna, S.K., Patterson, A.D., Yang, Q., Krausz, K.W., Ilde, J.R., Fornace Jr., A.J. & Gonzalez, F.J. 2011. UPLC–MS-based urine metabolomics reveals indole-3-lactic acid and phenyllactic acid as conserved biomarkers for alcohol-induced liver disease in the *Ppara*-null mouse model. *Journal of proteome research*, 10(9):4120–4133.

Mardh, G., Luehr, C.A. & Vallee, B.L. 1985. Human class I alcohol dehydrogenases catalyze the oxidation of glycols in the metabolism of norepinephrine. *Proceedings of the National Academy of Sciences USA*, 82(15):4979–4982.

Mitchell, S.C., Waring, R.H. & Wilson, I.D. 2014. Ethyl sulphate, a chemically reactive human metabolite of ethanol? *Xenobiotica*, 44(11):957–960.

- Oreland, L., Lagravinese, G., Toffoletto, S., Nilsson, K.W., Harro, J., Cloninger, C.R. & Comasco, E. 2017. Personality as an intermediate phenotype for genetic dissection of alcohol use disorder. *Journal of neural transmission*:1–24. doi:10.1007/s00702-016-1672-9.
- Peng, G-S., Wang M-F., Chen, C-Y., Luu, S.U., Chou, H.C., Li, T-K. & Yin, S-J. 1999. Involvement of acetaldehyde for full protection against alcoholism by homozygosity of the variant allele of mitochondrial aldehyde dehydrogenase gene in Asians. *Pharmacogenetics*, 9(4):463–476.
- Plantinga Jr., C. 1998. Pray the Lord my mind to keep. *Christianity today*, 42(9). <http://www.christianitytoday.com/ct/1998/august10/8t9050.html> Date of access: 26 Oct. 2017.
- Satoh, T. 2007. Genetic polymorphism in drug metabolism and toxicity: Linking animal research and risk assessment in man. *AATEX*, 14:443–445.
- Skipper, G.E., Weinmann, W., Thierauf, A., Schaefer, P., Wiesbeck, G., Pallen, J.P., Miller, M. & Wurst, F.M. 2004. Ethyl glucuronide: a biomarker to identify alcohol use by health professionals recovering from substance use disorders. *Alcohol & alcoholism*, 39(5):445–449.
- Smith, K. 2002. Genetic polymorphism and SNPs: genotyping, haplotype assembly problem, haplotype map, functional genomics and proteomics. *School of computer science, McGill University*. http://www.cs.mcgill.ca/~kaleigh/compbio/snp/snp_summary.html Date of access: 24 Apr. 2016.
- Wurst, F.M., Alling, C., Aradottir, S., Pragst, F., Allen, J.P., Weinmann, W., Marmillot, P., Ghosh, P., Lakshman, R., Skipper, G.E., Neumann, T., Spies, C., Javors, M., Johnson, B.A., Ait-Daoud, N., Akhtar, F., Roache, J.D. & Litten, R. 2005. Emerging biomarkers: new directions and clinical applications. *Alcoholism: clinical and experimental research*, 29(3):465–473.
- Zakhari, S. & Li, T-K. 2007. Determinants of alcohol use and abuse: impact of quantity and frequency patterns on liver disease. *Hepatology*, 46(6):2032–2039.
- Zintzaras, E., Stefanidis, I., Santos, M. & Vidal, F. 2006. Do alcohol metabolizing enzyme gene polymorphisms increase the risk of alcoholism and alcoholic liver disease? *Hepatology*, 43(2):352–361.

"Promise me you'll always remember: You're
braver than you believe, and stronger
than you seem, and smarter than you
think." - AA Milne

ANNEXURE

1 Ethical approval certificate



NORTH-WEST UNIVERSITY
YUNIBESITHI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

Private Bag X6001, Potchefstroom
South Africa 2520

Tel: (018) 299-4900
Faks: (018) 299-4910
Web: <http://www.nwu.ac.za>

Ethics Committee

Tel +27 18 299 4850
Fax +27 18 293 5329
Email Ethics@nwu.ac.za

ETHICS APPROVAL OF PROJECT

6 December 2012

The North-West University Ethics Committee (NWU-EC) hereby approves your project as indicated below. This implies that the NWU-EC grants its permission that, provided the special conditions specified below are met and pending any other authorisation that may be necessary, the project may be initiated, using the ethics number below.

Project title : An investigation into the metabolic responses to acute alcohol consumption

Project Leader Prof C.J Reinecke

N	W	U	-	0	0	0	4	5	-	1	2	-	A	1
Institution			Project Number							Year		Status		

Ethics number:

Status: S = Submission; R = Re-Submission; P = Provisional Authorisation; A = Authorisation

Approval date: 2012/12/06

Expiry date: 2017/12/05

Special conditions of the approval (if any): None

General conditions:

While this ethics approval is subject to all declarations, undertakings and agreements incorporated and signed in the application form, please note the following:

- The project leader (principal investigator) must report in the prescribed format to the NWU-EC:
 - annually (or as otherwise requested) on the progress of the project,
 - without any delay in case of any adverse event (or any matter that interrupts sound ethical principles) during the course of the project.
- The approval applies strictly to the protocol as stipulated in the application form. Would any changes to the protocol be deemed necessary during the course of the project, the project leader must apply for approval of these changes at the NWU-EC. Would there be deviation from the project protocol without the necessary approval of such changes, the ethics approval is immediately and automatically forfeited.
- The date of approval indicates the first date that the project may be started. Would the project have to continue after the expiry date, a new application must be made to the NWU-EC and new approval received before or on the expiry date.
- In the interest of ethical responsibility the NWU-EC retains the right to:
 - request access to any information or data at any time during the course or after completion of the project;
 - withdraw or postpone approval if:
 - any unethical principles or practices of the project are revealed or suspected,
 - it becomes apparent that any relevant information was withheld from the NWU-EC or that information has been false or misrepresented,
 - the required annual report and reporting of adverse events was not done timely and accurately,
 - new institutional rules, national legislation or international conventions deem it necessary.

The Ethics Committee would like to remain at your service as scientist and researcher, and wishes you well with your project. Please do not hesitate to contact the Ethics Committee for any further enquiries or requests for assistance.

Yours sincerely

Prof Amanda Lourens
(chair NWU Ethics Committee)

2 Informed consent form

INFORMED CONSENT TO PARTICIPATE IN EXPERIMENT WITH ETHICAL APPROVAL FROM THE NORTH-WESTUNIVERSITY

TITLE:

An investigation into the metabolic responses to acute alcohol consumption

AIM OF EXPERIMENT:

To determine the metabolic perturbations in young males in a fasted state, due to the consumption of a fixed acute dose of alcohol and/or NAD, as investigated by a metabolomics methodology.

INVESTIGATOR'S NAME: Cindy Irwin

SUPERVISOR: Prof. C.J. Reinecke

INVESTIGATOR SITE NAME & ADDRESS: Centre for Human Metabonomics
NWU, Potchefstroom Campus
Private Bag X6001
POTCHEFSTROOM 2520
South Africa
Tel: 018 299 2309
Fax: 018 293 5248

INTRODUCTION

The North-West University's Centre for Human Metabolomics aims to investigate perturbations associated with human metabolism by means of a metabolomics approach. Metabolomics is a biochemical technique, involving a comprehensive study of low molecular weight biomolecules, commonly known as metabolites. Biochemical analysis of biological samples (such as urine or blood) provides a large comprehensive list of metabolites. The data are analysed by means of bioinformatics, a field of science incorporating statistical multivariate analysis, providing information used to determine/distinguish any potential irregularities within the metabolic profile. The focus is the identification of very specific metabolites that can statistically discriminate between normal and abnormal metabolic situations. These metabolites are known as biomarkers.

PURPOSE OF THE EXPERIMENT

This experiment constitutes part of the investigator's M.Sc. thesis involving the study into the metabolic perturbations associated with alcohol consumption. This experiment is aimed at determining the changes in the metabolite profile that occur after the consumption of an acute alcohol dose and what, if any, effect the administration of NAD together with the alcohol has to prevent these metabolic changes. To minimize variation, a homogeneous, defined experimental group, namely young males between the ages of 20 and 30 years in an overnight fasted state, will be used. The biological specimens used in this experiment will be a blood sample taken before the start of the experiment, as well as urine samples taken at defined intervals of time, followed by a metabolomics analysis.

PROTOCOL OF EXPERIMENT

Acute Alcohol and/or NAD Dose Study:

- Participants are required to abstain from alcohol consumption at least 48 hours preceding this

experiment; as well as abstaining from consumption of food prior to initiation of the experiment (i.e. overnight fasted state).

- A questionnaire involving the basic clinical profile, including history of alcohol consumption, of the participant needs to be completed.
- An initial urine sample will be taken and labelled “0 hour” before alcohol and/or NAD dose is consumed.
- Group 1 participants will consume only the vehicle with which the alcohol will be mixed (i.e. lemon flavoured water) over a 15-minute period.
- Group 2 participants will consume a fixed dose of alcohol (Smirnoff vodka 43%, v/v, mixed with lemon flavoured water) of 1.5 g/kg body weight amount over a 15-minute period.
- Group 3 participants will consume the vehicle (i.e. lemon flavoured water), to which 50 mg of NAD has been added, over a 15-minute period.
- Group 4 participants will consume a fixed dose of alcohol (Smirnoff vodka 43%, v/v, mixed with lemon flavoured water) of 1.5 g/kg body weight, to which 50 mg of NAD has been added, over a 15-minute period.
- Urine and saliva for metabolomic investigations will be collected at an agreed time sequence over a period of 5 hours after alcohol and/or NAD dose first consumed, yielding 7 data points (consumption of water 15 minutes before a voiding is needed to facilitate obtaining urine samples).
- The experiment will be done in an environment compatible with clinical and ethical requirements
- A general practitioner will attend the early phases of the experiment (first 2 hours) and will be on call for the remaining part of the experiment.

NOTE: Urine samples must be labelled clearly and precisely and stored in a refrigerated environment and no additional alcohol should be consumed during the experimental time frame.

INFORMED CONSENT PROCEDURES

Participation in the project is fully voluntary. You are free to enquire about the experiment through the investigator and/or supervisor and, if agreeing to participate, the participant will be asked to sign this informed consent form. Should any participant request feedback on the outcomes of this study, such information can be made available to them.

It is required from all participants in this study to complete a questionnaire which provides information which is essential for the project. The participating physician will evaluate the information of all participants, and will approve their participation based on the information given in the questionnaire.

BENEFITS ASSOCIATED WITH THE STUDY

The Master’s study, to which this experiment will contribute, is a research project aimed at better understanding the biological relationship between chronic/acute alcohol consumption and metabolic perturbations within humans. As in all cases, improved knowledge of the normal physiology will eventually yield to a better understanding and treatment of any deviation from normal physiology. The outcome of this research will be used by the researcher for a M.Sc. thesis and no reference will be included in the thesis regarding any individual who participated in the study.

PAYMENT OR REIMBURSEMENT

Participants will not be paid for their participation and do not contribute to the costs of the study. An amount of R100 will however be paid to each of the participants for travel expenses and other inconveniences that resulted from participation in the study as well as R50 for the light meal of choice after completion of the experiment.

CONFIDENTIALITY

All research records are confidential unless the law requires disclosure. No name or other personal identifying information of the participants will be used in any reports or publications resulting from this study. Data from this study will be used in an anonymous statistical analysis and reported as such by the NWU. No patient's identification details will be reported or made known to other parties.

VOLUNTARY PARTICIPATION AND CONDITIONS OF WITHDRAWAL

Your participation in this study is completely voluntary. You may choose not to participate in this study to which you are otherwise entitled.

CONSENT

I, _____, have read and understood the preceding information describing this research study and my questions have been answered to my satisfaction. I voluntarily consent to participate in this research study. I do not waive my legal rights by signing this consent form. I will receive a signed and dated copy of this consent form.

PARTICIPANT:

_____	_____	_____
Printed name	Signature	Date

INVESTIGATOR:

_____	_____	_____
Printed name	Signature	Date

PHYSICIAN:

_____	_____	_____
Printed name	Signature	Date

3 Copyright licences of journals

3.1 PLOS ONE (1)

Contribution towards a metabolite profile of the detoxification of benzoic acid through glycine conjugation: an intervention study (Irwin *et al.*, 2016)

Copyright: © 2016 Irwin *et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

3.2 PLOS ONE (2)

The ¹H-NMR-based metabolite profile of acute alcohol consumption: a metabolomics intervention study

This paper has been accepted for publication in *PLOS ONE*, but has not yet been published. Also, PLOS applies the Creative Commons Attribution (CC BY) license to works they publish. This license was developed to facilitate open access – namely, free immediate access to, and unrestricted reuse of, original works of all types. Under this license, authors agree to make articles legally available for reuse, without permission or fees, for virtually any purpose. Anyone may copy, distribute, or reuse these articles, as long as the author and original source are properly cited. For these reasons no copyright licence is required.

3.3 Scientific Reports

GC–MS-based urinary organic acid profiling reveals multiple dysregulated metabolic pathways following experimental acute alcohol consumption (Irwin *et al.*, 2018)

Copyright: © The Author(s) 2018. This is an open access article licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the

copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

4 Instructions to authors from journals

4.1 PLOS ONE



Journal Information

The world's first multidisciplinary Open Access journal, *PLOS ONE* accepts scientifically rigorous research, regardless of novelty. *PLOS ONE*'s broad scope provides a platform to publish primary research, including interdisciplinary and replication studies as well as negative results. The journal's publication criteria are based on high ethical standards and the rigor of the methodology and conclusions reported.

Scope

PLOS ONE features reports of original research from all disciplines within science and medicine. By not excluding research on the basis of subject area, *PLOS ONE* facilitates the discovery of connections between research whether within or between disciplines.

We will also consider the following article types:

- **Systematic reviews.** We consider publishing systematic reviews only if the methods ensure the comprehensive and unbiased sampling of existing literature.
- **Submissions describing methods, software, databases, or other tools.** We consider submissions describing methods, software, databases, or other tools if they follow the appropriate reporting guidelines.
- **Qualitative research.** We consider publishing qualitative research only if it adheres to appropriate study design and reporting guidelines.
- **Studies reporting negative results.**

Criteria for Publication:

1. The study presents the results of primary scientific research.
2. Results reported have not been published elsewhere.
3. Experiments, statistics, and other analyses are performed to a high technical standard and are described in sufficient detail.
4. Conclusions are presented in an appropriate fashion and are supported by the data.
5. The article is presented in an intelligible fashion and is written in standard English.
6. The research meets all applicable standards for the ethics of experimentation and research integrity.
7. The article adheres to appropriate reporting guidelines and community standards for data availability.

Rigorous Peer Review

Often a journal's decision not to publish a paper reflects an editor's opinion about what is likely to have substantial impact in a given field. These subjective judgments can delay the publication of work that later proves to be of major significance. *PLOS ONE* will rigorously peer-review your submissions and publish all papers that are judged to be technically sound. Judgments about the importance of any particular paper are then made after publication by the readership, who are the most qualified to determine what is of interest to them.

Each submission to *PLOS ONE* passes through a rigorous quality control and peer-review evaluation process before receiving a decision. The initial in-house quality control check deals with issues such as competing interests; ethical requirements for studies involving human participants or animals; financial disclosures; full compliance with PLOS'

data availability policy, etc. Submissions may be returned to authors for queries, and will not be seen by our Editorial Board or peer reviewers until they pass this quality control check.

Once each manuscript has passed quality control, it is assigned to a member of the Editorial Board, who takes responsibility as the Academic Editor for the submission. The Academic Editor is responsible for conducting the peer-review process and for making a decision to accept, invite revision of, or reject the article.

Submission Guidelines

Style and Format

File format	Manuscript files can be in the following formats: DOC, DOCX, RTF, or PDF. Microsoft Word documents should not be locked or protected. LaTeX manuscripts must be submitted as PDFs. Read the LaTeX guidelines.						
Length	Manuscripts can be any length. There are no restrictions on word count, number of figures, or amount of supporting information. We encourage you to present and discuss your findings concisely.						
Font	Use a standard font size and any standard font, except for the font named “Symbol”. To add symbols to the manuscript, use the Insert → Symbol function in your word processor or paste in the appropriate Unicode character.						
Headings	Limit manuscript sections and sub-sections to 3 heading levels. Make sure heading levels are clearly indicated in the manuscript text.						
Layout and spacing	Manuscript text should be double-spaced. Do not format text in multiple columns.						
Page and line numbers	Include page numbers and line numbers in the manuscript file. Use continuous line numbers (do not restart the numbering on each page).						
Footnotes	Footnotes are not permitted. If your manuscript contains footnotes, move the information into the main text or the reference list, depending on the content.						
Language	Manuscripts must be submitted in English. You may submit translations of the manuscript or abstract as supporting information. Read the supporting information guidelines.						
Abbreviations	Define abbreviations upon first appearance in the text. Do not use non-standard abbreviations unless they appear at least three times in the text. Keep abbreviations to a minimum.						
Reference style	PLOS uses “Vancouver” style, as outlined in the ICMJE sample references. See reference formatting examples and additional instructions below.						
Equations	We recommend using MathType for display and inline equations, as it will provide the most reliable outcome. If this is not possible, Equation Editor or Microsoft's Insert→Equation function is acceptable. Avoid using MathType, Equation Editor, or the Insert→Equation function to insert single variables (e.g., “ $a^2 + b^2 = c^2$ ”), Greek or other symbols (e.g., β , Δ , or ' [prime]), or mathematical operators (e.g., $\times$, $\geq$, or $\pm$) in running text. Wherever possible, insert single symbols as normal text with the correct Unicode (hex) values. Do not use MathType, Equation Editor, or the Insert→Equation function for only a portion of an equation. Rather, ensure that the entire equation is included. Equations should not contain a mix of different equation tools. Avoid “hybrid” inline or display equations, in which part is text and part is MathType, or part is MathType and part is Equation Editor.						
Nomenclature	Use correct and established nomenclature wherever possible. <table> <tr> <td><i>Units of measurement</i></td><td>Use SI units. If you do not use these exclusively, provide the SI value in parentheses after each value. Read more about SI units.</td></tr> <tr> <td><i>Drugs</i></td><td>Provide the Recommended International Non-Proprietary Name (rINN).</td></tr> <tr> <td><i>Species names</i></td><td>Write in italics (e.g., <i>Homo sapiens</i>). Write out in full the genus and species, both in the title of the manuscript and at the first mention</td></tr> </table>	<i>Units of measurement</i>	Use SI units. If you do not use these exclusively, provide the SI value in parentheses after each value. Read more about SI units.	<i>Drugs</i>	Provide the Recommended International Non-Proprietary Name (rINN).	<i>Species names</i>	Write in italics (e.g., <i>Homo sapiens</i>). Write out in full the genus and species, both in the title of the manuscript and at the first mention
<i>Units of measurement</i>	Use SI units. If you do not use these exclusively, provide the SI value in parentheses after each value. Read more about SI units.						
<i>Drugs</i>	Provide the Recommended International Non-Proprietary Name (rINN).						
<i>Species names</i>	Write in italics (e.g., <i>Homo sapiens</i>). Write out in full the genus and species, both in the title of the manuscript and at the first mention						

	of an organism in a paper. After first mention, the first letter of the genus name followed by the full species name may be used (e.g., <i>H. sapiens</i>). Write in italics. Use the recommended name by consulting the appropriate genetic nomenclature database (e.g., HUGO for human genes). It is sometimes advisable to indicate the synonyms for the gene the first time it appears in the text. Gene prefixes such as those used for oncogenes or cellular localization should be shown in roman typeface (e.g., v-fes, c-MYC). The systematic allergen nomenclature of the World Health Organization/International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Sub-committee should be used for manuscripts that include the description or use of allergenic proteins. For manuscripts describing new allergens, the systematic name of the allergen should be approved by the WHO/IUIS Allergen Nomenclature Sub-Committee prior to manuscript publication. Examples of the systematic allergen nomenclature can be found at the WHO/IUIS Allergen Nomenclature site.
<i>Genes, mutations, genotypes, and alleles</i>	
<i>Allergens</i>	

Copyediting manuscripts

Prior to submission, authors who believe their manuscripts would benefit from professional editing are encouraged to use language-editing and copyediting services. Obtaining this service is the responsibility of the author, and should be done before initial submission. These services can be found on the web using search terms like “scientific editing service” or “manuscript editing service.”

Submissions are not copyedited before publication.

Submissions that do not meet the [PLOS ONE publication criterion for language standards](#) may be rejected.

Manuscript Organization

Manuscripts should be organized as follows. Instructions for each element appear below the list.

Beginning section	<i>The following elements are required, in order:</i> • Title page: List title, authors, and affiliations as first page of manuscript • Abstract • Introduction
Middle section	<i>The following elements can be renamed as needed and presented in any order:</i> • Materials and Methods • Results • Discussion • Conclusions (optional)
Ending section	<i>The following elements are required, in order:</i> • Acknowledgments • References • Supporting information captions (if applicable)
Other elements	• Figure captions are inserted immediately after the first paragraph in which the figure is cited. Figure files are uploaded separately. • Tables are inserted immediately after the first paragraph in which they are cited. • Supporting information files are uploaded separately.

Viewing Figures and Supporting Information in the compiled submission PDF

The compiled submission PDF includes low-resolution preview images of the figures after the reference list. The function of these previews is to allow you to download the entire submission as quickly as possible. Click the link at the top of each preview page to download a high-resolution version of each figure. Links to download Supporting Information files are also available after the reference list.

Parts of a Submission

Title

Include a full title and a short title for the manuscript.

Title	Length	Guidelines	Examples
Full title	250 characters	Specific, descriptive, concise, and comprehensible to readers outside the field	Impact of cigarette smoke exposure on innate immunity: A <i>Caenorhabditis elegans</i> model Solar drinking water disinfection (SODIS) to reduce childhood diarrhoea in rural Bolivia: A cluster-randomized, controlled trial
Short title	100 characters	State the topic of the study	Cigarette smoke exposure and innate immunity SODIS and childhood diarrhoea

Titles should be written in sentence case (only the first word of the text, proper nouns, and genus names are capitalized). Avoid specialist abbreviations if possible. For clinical trials, systematic reviews, or meta-analyses, the subtitle should include the study design.

Author list

Author names and affiliations

Enter author names on the title page of the manuscript and in the online submission system.

On the title page, write author names in the following order:

- First name (or initials, if used)
- Middle name (or initials, if used)
- Last name (surname, family name)

Each author on the list must have an affiliation. The affiliation includes department, university, or organizational affiliation and its location, including city, state/province (if applicable), and country. Authors have the option to include a current address in addition to the address of their affiliation at the time of the study. The current address should be listed in the byline and clearly labeled “current address.” At a minimum, the address must include the author’s current institution, city, and country.

If an author has multiple affiliations, enter all affiliations on the title page only. In the submission system, enter only the preferred or primary affiliation. Author affiliations will be listed in the typeset PDF article in the same order that authors are listed in the submission.

Corresponding author

The submitting author is automatically designated as the corresponding author in the submission system. The corresponding author is the primary contact for the journal office and the only author able to view or change the manuscript while it is under editorial consideration.

The corresponding author role may be transferred to another coauthor. However, note that transferring the corresponding author role also transfers access to the manuscript. (To designate a new corresponding author while the manuscript is still under consideration, watch the video tutorial below.)

Only one corresponding author can be designated in the submission system, but this does not restrict the number of corresponding authors that may be listed on the article in the event of publication. Whoever is designated as a corresponding author on the title page of the manuscript file will be listed as such upon publication. Include an email address for each corresponding author listed on the title page of the manuscript.

Author contributions

Provide at minimum one contribution for each author in the submission system. Use the CRediT taxonomy to describe each contribution. [Read the policy and the full list of roles](#).

Contributions will be published with the final article, and they should accurately reflect contributions to the work. The submitting author is responsible for completing this information at submission, and we expect that all authors will have reviewed, discussed, and agreed to their individual contributions ahead of this time.

PLOS ONE will contact all authors by email at submission to ensure that they are aware of the submission.

Cover letter

Upload a cover letter as a separate file in the online system. The length limit is 1 page.

The cover letter should include the following information:

- Summarize the study's contribution to the scientific literature
- Relate the study to previously published work
- Specify the type of article (for example, research article, systematic review, meta-analysis, clinical trial)
- Describe any prior interactions with PLOS regarding the submitted manuscript
- Suggest appropriate Academic Editors to handle your manuscript ([see the full list of Academic Editors](#))
- List any opposed reviewers

Title page

The title, authors, and affiliations should all be included on a title page as the first page of the manuscript file.

Abstract

The Abstract comes after the title page in the manuscript file. The abstract text is also entered in a separate field in the submission system.

The Abstract should:

- Describe the main objective(s) of the study
- Explain how the study was done, including any model organisms used, without methodological detail
- Summarize the most important results and their significance
- Not exceed 300 words

Abstracts should not include:

- Citations
- Abbreviations, if possible

Introduction

The introduction should:

- Provide background that puts the manuscript into context and allows readers outside the field to understand the purpose and significance of the study
- Define the problem addressed and why it is important
- Include a brief review of the key literature
- Note any relevant controversies or disagreements in the field
- Conclude with a brief statement of the overall aim of the work and a comment about whether that aim was achieved

Materials and Methods

The Materials and Methods section should provide enough detail to allow suitably skilled investigators to fully replicate your study. Specific information and/or protocols for new methods should be included in detail. If materials, methods, and protocols are well established, authors may cite articles where those protocols are described in detail, but the submission should include sufficient information to be understood independent of these references.

Protocol documents for clinical trials, observational studies, and other **non-laboratory** investigations may be uploaded as supporting information. [Read the supporting information guidelines](#) for formatting instructions. We recommend depositing **laboratory protocols** at [protocols.io](#). Read detailed [instructions for depositing and sharing your laboratory protocols](#).

Human or animal subjects and/or tissue or field sampling

Methods sections describing research using human or animal subjects and/or tissue or field sampling must include required ethics statements. See the [reporting guidelines for human research, clinical trials, animal research, and observational and field studies](#) for more information.

Data

PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.

Large data sets, including raw data, may be deposited in an appropriate public repository. [See our list of recommended repositories](#).

For smaller data sets and certain data types, authors may provide their data within [supporting information files](#) accompanying the manuscript. Authors should take care to maximize the accessibility and reusability of the data by selecting a file format from which data can be efficiently extracted (for example, spreadsheets or flat files should be provided rather than PDFs when providing tabulated data).

For more information on how best to provide data, read our [policy on data availability](#). PLOS does not accept references to “data not shown.”

Cell lines

Methods sections describing research using cell lines must state the origin of the cell lines used. [See the reporting guidelines for cell line research for more information](#).

Laboratory Protocols

To enhance the reproducibility of your results, we recommend and encourage you to deposit laboratory protocols in [protocols.io](#), where protocols can be assigned their own persistent digital object identifiers (DOIs).

To include a link to a protocol in your article:

1. Describe your step-by-step protocol on protocols.io
2. Select **Get DOI** to issue your protocol a persistent digital object identifier (DOI)

3. Include the DOI link in the Methods section of your manuscript using the following format provided by protocols.io: [http://dx.doi.org/10.17504/protocols.io.\[PROTOCOL DOI\]](http://dx.doi.org/10.17504/protocols.io.[PROTOCOL DOI])

At this stage, your protocol is only visible to those with the link. This allows editors and reviewers to consult your protocol when evaluating the manuscript. You can make your protocols public at any time by selecting **Publish** on the protocols.io site. Any referenced protocol(s) will automatically be made public when your article is published.

New taxon names

Methods sections of manuscripts adding new taxon names to the literature must follow the [reporting guidelines below for a new zoological taxon, botanical taxon, or fungal taxon](#).

Results, Discussion, Conclusions

These sections may all be separate, or may be combined to create a mixed Results/Discussion section (commonly labeled “Results and Discussion”) or a mixed Discussion/Conclusions section (commonly labeled “Discussion”). These sections may be further divided into subsections, each with a concise subheading, as appropriate. These sections have no word limit, but the language should be clear and concise.

Together, these sections should describe the results of the experiments, the interpretation of these results, and the conclusions that can be drawn.

Authors should explain how the results relate to the hypothesis presented as the basis of the study and provide a succinct explanation of the implications of the findings, particularly in relation to previous related studies and potential future directions for research.

PLOS ONE editorial decisions do not rely on perceived significance or impact, so authors should avoid overstating their conclusions. See the [PLOS ONE Criteria for Publication](#) for more information.

Acknowledgments

Those who contributed to the work but do not meet our authorship criteria should be listed in the Acknowledgments with a description of the contribution.

Authors are responsible for ensuring that anyone named in the Acknowledgments agrees to be named.

References

Any and all available works can be cited in the reference list. Acceptable sources include:

- Published or accepted manuscripts
- Manuscripts on preprint servers, if the manuscript is submitted to a journal and also publicly available as a preprint

Do not cite the following sources in the reference list:

- Unavailable and unpublished work, including manuscripts that have been submitted but not yet accepted (e.g., “unpublished work,” “data not shown”). Instead, include those data as supplementary material or deposit the data in a publicly available database.
- Personal communications (these should be supported by a letter from the relevant authors but not included in the reference list)

References are listed at the end of the manuscript and numbered in the order that they appear in the text. In the text, cite the reference number in square brackets (e.g., “We used the techniques developed by our colleagues [19] to analyze the data”). *PLOS* uses the numbered citation (citation-sequence) method and first six authors, et al.

Do not include citations in abstracts or author summaries.

Make sure the parts of the manuscript are in the correct order *before* ordering the citations.

Formatting references

PLOS uses the reference style outlined by the International Committee of Medical Journal Editors (ICMJE), also referred to as the “Vancouver” style. Example formats are listed below. Additional examples are in the [ICMJE sample references](#).

A reference management tool, EndNote, offers a current [style file](#) that can assist you with the formatting of your references. If you have problems with any reference management program, please contact the source company's technical support.

Journal name abbreviations should be those found in the [National Center for Biotechnology Information \(NCBI\) databases](#).

Source	Format
Published articles	Hou WR, Hou YL, Wu GF, Song Y, Su XL, Sun B, et al. cDNA, genomic sequence cloning and overexpression of ribosomal protein gene L9 (rpL9) of the giant panda (<i>Ailuropoda melanoleuca</i>). Genet Mol Res. 2011;10: 1576-1588. Devaraju P, Gulati R, Antony PT, Mithun CB, Negi VS. Susceptibility to SLE in South Indian Tamils may be influenced by genetic selection pressure on TLR2 and TLR9 genes. Mol Immunol. 2014 Nov 22. pii: S0161-5890(14)00313-7. doi: 10.1016/j.molimm.2014.11.005. <i>Note: A DOI number for the full-text article is acceptable as an alternative to or in addition to traditional volume and page numbers.</i>
Accepted, unpublished articles	Same as published articles, but substitute “Forthcoming” for page numbers or DOI.
Web sites or online articles	Huynen MMTE, Martens P, Hilderlink HBM. The health impacts of globalisation: a conceptual framework. Global Health. 2005;1: 14. Available from: http://www.globalizationandhealth.com/content/1/1/14
Books	Bates B. Bargaining for life: A social history of tuberculosis. 1st ed. Philadelphia: University of Pennsylvania Press; 1992.
Book chapters	Hansen B. New York City epidemics and history for the public. In: Harden VA, Risse GB, editors. AIDS and the historian. Bethesda: National Institutes of Health; 1991. pp. 21-28.
Deposited articles (preprints, e-prints, or arXiv)	Krick T, Shub DA, Verstraete N, Ferreira DU, Alonso LG, Shub M, et al. Amino acid metabolism conflicts with protein diversity; 1991. Preprint. Available from: arXiv:1403.3301v1 . Cited 17 March 2014.
Published media (print or online newspapers and magazine articles)	Fountain H. For Already Vulnerable Penguins, Study Finds Climate Change Is Another Danger. The New York Times. 29 Jan 2014. Available from: http://www.nytimes.com/2014/01/30/science/earth/climate-change-taking-toll-on-penguins-study-finds.html Cited 17 March 2014.
New media (blogs, web sites, or other written works)	Allen L. Announcing PLOS Blogs. 2010 Sep 1 [cited 17 March 2014]. In: PLOS Blogs [Internet]. San Francisco: PLOS 2006 - . [about 2 screens]. Available from: http://blogs.plos.org/plos/2010/09/announcing-plos-blogs/ .
Masters' theses or doctoral dissertations	Wells A. Exploring the development of the independent, electronic, scholarly journal. M.Sc. Thesis, The University of Sheffield. 1999. Available from: http://cumincad.scix.net/cgi-bin/works/Show?2e09
Databases and repositories (Figshare, arXiv)	Roberts SB. QPX Genome Browser Feature Tracks; 2013 [cited 2013 Oct 5]. Database: figshare [Internet]. Available from: http://figshare.com/articles/QPX_Genome_Browser_Feature_Tracks/701214
Multimedia (videos, movies, or TV shows)	Hitchcock A, producer and director. Rear Window [Film]; 1954. Los Angeles: MGM.

Supporting Information

Authors can submit essential supporting files and multimedia files along with their manuscripts. All supporting information will be subject to peer review. All file types can be submitted, but files must be smaller than 10 MB in size.

Authors may use almost any description as the item name for a supporting information file as long as it contains an “S” and number. For example, “S1 Appendix” and “S2 Appendix,” “S1 Table” and “S2 Table,” and so forth.

Supporting information files are published exactly as provided, and are not copyedited.

Supporting information captions

List supporting information captions at the end of the manuscript file. Do not submit captions in a separate file.

The file number and name are required in a caption, and we highly recommend including a one-line title as well. You may also include a legend in your caption, but it is not required.

Example caption

S1 Text. Title is strongly recommended. Legend is optional.

In-text citations

We recommend that you cite supporting information in the manuscript text, but this is not a requirement. If you cite supporting information in the text, citations do not need to be in numerical order.

Read the [supporting information guidelines](#) for more details about submitting supporting information and multimedia files.

Figures and Tables

Figures

Do not include figures in the main manuscript file. Each figure must be prepared and submitted as an individual file.

Cite figures in ascending numeric order upon first appearance in the manuscript file.

[Read the guidelines for figures.](#)

Figure captions

Figure captions must be inserted in the text of the manuscript, immediately following the paragraph in which the figure is first cited (read order). Do not include captions as part of the figure files themselves or submit them in a separate document.

At a minimum, include the following in your figure captions:

- A figure label with Arabic numerals, and “Figure” abbreviated to “Fig” (e.g. Fig 1, Fig 2, Fig 3, etc). Match the label of your figure with the name of the file uploaded at submission (e.g. a figure citation of “Fig 1” must refer to a figure file named “Fig1.tif”).
- A concise, descriptive title

The caption may also include a legend as needed.

[Read more about figure captions.](#)

Tables

Cite tables in ascending numeric order upon first appearance in the manuscript file.

Place each table in your manuscript file directly after the paragraph in which it is first cited (read order). Do not submit your tables in separate files.

Tables require a label (e.g., “Table 1”) and brief descriptive title to be placed above the table. Place legends, footnotes, and other text below the table.

[Read the guidelines for tables.](#)

Data reporting

All data and related metadata underlying the findings reported in a submitted manuscript should be deposited in an appropriate public repository, unless already provided as part of the submitted article.

[Read our policy on data availability.](#)

Repositories may be either subject-specific (where these exist) and accept specific types of structured data, or generalist repositories that accept multiple data types. We recommend that authors select repositories appropriate to their field. Repositories may be subject-specific (e.g., GenBank for sequences and PDB for structures), general, or institutional, as long as DOIs or accession numbers are provided and the data are at least as open as CC BY. Authors are encouraged to select repositories that meet accepted criteria as trustworthy digital repositories, such as criteria of the Centre for Research Libraries or Data Seal of Approval. Large, international databases are more likely to persist than small, local ones.

[See our list of recommended repositories.](#)

To support data sharing and author compliance of the PLOS data policy, we have integrated our submission process with a select set of data repositories. The list is neither representative nor exhaustive of the suitable repositories available to authors. Current repository integration partners include [Dryad](#) and [FlowRepository](#). Please contact data@plos.org to make recommendations for further partnerships.

Instructions for PLOS submissions with data deposited in an integration partner repository:

- Deposit data in the integrated repository of choice.
- Once deposition is final and complete, the repository will provide you with a dataset DOI (provisional) and private URL for reviewers to gain access to the data.
- Enter the given data DOI into the full Data Availability Statement, which is requested in the Additional Information section of the PLOS submission form. Then provide the URL passcode in the Attach Files section.

If you have any questions, please [email us](#).

Accession numbers

All appropriate data sets, images, and information should be deposited in an appropriate public repository. [See our list of recommended repositories.](#)

Accession numbers (and version numbers, if appropriate) should be provided in the Data Availability Statement. Accession numbers or a citation to the DOI should also be provided when the data set is mentioned within the manuscript.

In some cases authors may not be able to obtain accession numbers of DOIs until the manuscript is accepted; in these cases, the authors must provide these numbers at acceptance. In all other cases, these numbers must be provided at submission.

Identifiers

As much as possible, please provide accession numbers or identifiers for all entities such as genes, proteins, mutants, diseases, etc., for which there is an entry in a public database, for example:

- [Ensembl](#)
- [Entrez Gene](#)
- [FlyBase](#)
- [InterPro](#)
- [Mouse Genome Database \(MGD\)](#)
- [Online Mendelian Inheritance in Man \(OMIM\)](#)
- [PubChem](#)

Identifiers should be provided in parentheses after the entity on first use.

Striking image

You can choose to upload a “Striking Image” that we may use to represent your article online in places like the journal homepage or in search results.

The striking image must be derived from a figure or supporting information file from the submission, i.e., a cropped portion of an image or the entire image. Striking images should ideally be high resolution, eye-catching, single panel images, and should ideally avoid containing added details such as text, scale bars, and arrows.

If no striking image is uploaded, we will designate a figure from the submission as the striking image.

Striking images should not contain potentially identifying images of people. [Read our policy on identifying information.](#)

[The PLOS licenses and copyright policy](#) also applies to striking images.

Additional Information Requested at Submission

Funding Statement

This information should not be in your manuscript file; you will provide it via our submission system.

This information will be published with the final manuscript, if accepted, so please make sure that this is accurate and as detailed as possible. You should not include this information in your manuscript file, but it is important to gather it prior to submission, because your financial disclosure statement cannot be changed after initial submission.

Your statement should include relevant grant numbers and the URL of any funder's web site. Please also state whether any individuals employed or contracted by the funders (other than the named authors) played any role in: study design, data collection and analysis, decision to publish, or preparation of the manuscript. If so, please name the individual and describe their role.

[Read our policy on disclosure of funding sources.](#)

Competing Interests

This information should not be in your manuscript file; you will provide it via our submission system.

All potential competing interests must be declared in full. If the submission is related to any patents, patent applications, or products in development or for market, these details, including patent numbers and titles, must be disclosed in full.

[Read our policy on competing interests.](#)

Manuscripts disputing published work

For manuscripts disputing previously published work, it is *PLOS ONE* policy to invite a signed review by the disputed author during the peer review process. This procedure is aimed at ensuring a thorough, transparent, and productive review process.

If the disputed author chooses to submit a review, it must be returned in a timely fashion and contain a full declaration of all competing interests. The Academic Editor will consider any such reviews in light of the competing interest.

Authors submitting manuscripts disputing previous work should explain the relationship between the manuscripts in their cover letter, and will be required to confirm that they accept the conditions of this review policy before the manuscript is considered further.

Related manuscripts

Upon submission, authors must confirm that the manuscript, or any related manuscript, is not currently under consideration or accepted elsewhere. If related work has been submitted to *PLOS ONE* or elsewhere, authors must include a copy with the submitted article. Reviewers will be asked to comment on the overlap between related submissions.

We strongly discourage the unnecessary division of related work into separate manuscripts, and we will not consider manuscripts that are divided into “parts.” Each submission to *PLOS ONE* must be written as an independent unit and should not rely on any work that has not already been accepted for publication. If related manuscripts are submitted to *PLOS ONE*, the authors may be advised to combine them into a single manuscript at the editor's discretion.

PLOS does support authors who wish to share their work early and receive feedback before formal peer review. Deposition of manuscripts with preprint servers does not impact consideration of the manuscript at any PLOS journal.

Authors choosing bioRxiv may now concurrently submit directly to select PLOS journals through [bioRxiv's direct transfer to journal service](#).

Read our policies on [related manuscripts](#) and [preprint servers](#).

Guidelines for Specific Study Types

Human subjects research

All research involving human participants must have been approved by the authors' Institutional Review Board (IRB) or by equivalent ethics committee(s), and must have been conducted according to the principles expressed in the [Declaration of Helsinki](#). Authors should be able to submit, upon request, a statement from the IRB or ethics committee indicating approval of the research. We reserve the right to reject work that we believe has not been conducted to a high ethical standard, even when formal approval has been obtained.

Subjects must have been properly instructed and have indicated that they consent to participate by signing the appropriate informed consent paperwork. Authors may be asked to submit a blank, sample copy of a subject consent form. If consent was verbal instead of written, or if consent could not be obtained, the authors must explain the reason in the manuscript, and the use of verbal consent or the lack of consent must have been approved by the IRB or ethics committee.

All efforts should be made to protect patient privacy and anonymity. Identifying information, including photos, should not be included in the manuscript unless the information is crucial and the individual has provided written consent by completing the [Consent Form for Publication in a PLOS Journal \(PDF\)](#). Download additional translations of the form from the [Downloads and Translations page](#). More information about patient privacy, anonymity, and informed consent can be found in the [International Committee of Medical Journal Editors \(ICMJE\) Privacy and Confidentiality guidelines](#).

Manuscripts should conform to the following reporting guidelines:

- Studies of diagnostic accuracy: [STARD](#)
- Observational studies: [STROBE](#)
- Microarray experiments: [MIAME](#)
- Other types of health-related research: Consult the [EQUATOR](#) web site for appropriate reporting guidelines

Methods sections of papers on research using human subjects or samples must include ethics statements that specify:

- **The name of the approving institutional review board or equivalent committee(s).** If approval was not obtained, the authors must provide a detailed statement explaining why it was not needed
- **Whether informed consent was written or oral.** If informed consent was oral, it must be stated in the manuscript:
 - Why written consent could not be obtained
 - That the Institutional Review Board (IRB) approved use of oral consent
 - How oral consent was documented

For studies involving humans categorized by race/ethnicity, age, disease/disabilities, religion, sex/gender, sexual orientation, or other socially constructed groupings, authors should:

- Explicitly describe their methods of categorizing human populations
- Define categories in as much detail as the study protocol allows
- Justify their choices of definitions and categories, including for example whether any rules of human categorization were required by their funding agency
- Explain whether (and if so, how) they controlled for confounding variables such as socioeconomic status, nutrition, environmental exposures, or similar factors in their analysis

In addition, outmoded terms and potentially stigmatizing labels should be changed to more current, acceptable terminology. Examples: “Caucasian” should be changed to “white” or “of [Western] European descent” (as appropriate); “cancer victims” should be changed to “patients with cancer.”

For papers that include identifying, or potentially identifying, information, authors must [download the Consent Form for Publication in a PLOS Journal](#), which the individual, parent, or guardian must sign once they have read the paper and been informed about the terms of PLOS open-access license. The signed consent form should not be submitted with the manuscript, but authors should securely file it in the individual's case notes and the methods section of the manuscript should explicitly state that consent authorization for publication is on file, using wording like:

The individual in this manuscript has given written informed consent (as outlined in PLOS consent form) to publish these case details.

For more information about *PLOS ONE* policies regarding human subjects research, see the [Publication Criteria](#) and [Editorial Policies](#).

Clinical trials

Clinical trials are subject to all [policies regarding human research](#). *PLOS ONE* follows the [World Health Organization's \(WHO\) definition of a clinical trial](#):

A clinical trial is any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes [...] Interventions include but are not restricted to drugs, cells and other biological products, surgical procedures, radiologic procedures, devices, behavioural treatments, process-of-care changes, preventive care, etc.

All clinical trials must be registered in one of the publicly-accessible registries approved by the [WHO](#) or [ICMJE](#) (International Committee of Medical Journal Editors). Authors must provide the trial registration number. Prior disclosure of results on a clinical trial registry site will not affect consideration for publication. We reserve the right to inform authors' institutions or ethics committees, and to reject the manuscript, if we become aware of unregistered trials.

PLOS ONE supports prospective trial registration (i.e. before participant recruitment has begun) as recommended by the ICMJE's [clinical trial registration policy](#). **Where trials were not publicly registered before participant recruitment began**, authors must:

- Register all related clinical trials and confirm they have done so in the Methods section
- Explain in the Methods the reason for failing to register before participant recruitment

Clinical trials must be reported according to the relevant reporting guidelines, i.e. [CONSORT](#) for randomized controlled trials, [TREND](#) for non-randomized trials, and [other specialized guidelines](#) as appropriate. The intervention should be described according to the requirements of the [TIDieR checklist and guide](#). Submissions must also include the study protocol as supporting information, which will be published with the manuscript if accepted.

Authors of manuscripts describing the results of clinical trials must adhere to the [CONSORT](#) reporting guidelines appropriate to their trial design, available on the [CONSORT Statement web site](#). Before the paper can enter peer review, authors must:

- Provide the registry name and number in the methods section of the manuscript

- Provide a copy of the trial protocol as approved by the ethics committee and a completed [CONSORT checklist](#) as supporting information (which will be published alongside the paper, if accepted). This should be named S1 CONSORT Checklist.
- Include the [CONSORT flow diagram](#) as the manuscript's "Fig 1"

Any deviation from the trial protocol must be explained in the paper. Authors must explicitly discuss informed consent in their paper, and we reserve the right to ask for a copy of the patient consent form.

The methods section must include the name of the registry, the registry number, and the URL of your trial in the registry database for each location in which the trial is registered.

Animal research

We work in consultation with the [PLOS ONE Animal Research Advisory Group](#) to develop policies. Animal Research Advisory Group members may also be consulted on individual submissions.

All research involving vertebrates or cephalopods must have approval from the authors' Institutional Animal Care and Use Committee (IACUC) or equivalent ethics committee(s), and must have been conducted according to applicable national and international guidelines. Approval must be received prior to beginning research.

If we note differences between an IACUC-approved protocol and the methods reported in a submitted manuscript, we may report these discrepancies to the relevant institution or committee.

Methods sections of manuscripts reporting results of animal research must include required ethics statements that specify:

- The full name of the relevant ethics committee that approved the work, and the associated permit number(s). Where ethical approval is not required, the manuscript should include a clear statement of this and the reason why.
- Relevant details for efforts taken to ameliorate animal suffering

Example ethics statement

This study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Committee on the Ethics of Animal Experiments of the University of Minnesota (Permit Number: 27-2956). All surgery was performed under sodium pentobarbital anesthesia, and all efforts were made to minimize suffering.

The organism(s) studied should always be stated in the abstract. Where research may be confused as pertaining to clinical research, the animal model should also be stated in the title.

Where unregulated animals are used or ethics approval is not required, authors should make this clear in submitted articles and explain why ethical approval was not required. Relevant regulations that grant exemptions should be cited in full. It is the authors' responsibility to understand and comply with all relevant regulations.

We reserve the right to reject work that the editors believe has not been conducted to a high ethical standard, even if authors have obtained formal approval or approval is not required under local regulations.

We encourage authors to follow the [Animal Research: Reporting of In Vivo Experiments \(ARRIVE\) guidelines](#) for all submissions describing laboratory-based animal research and to upload a completed [ARRIVE Guidelines Checklist](#) to be published as supporting information. Please note that inclusion of a completed ARRIVE Checklist may be a formal requirement for publication at a later date.

Non-human primates

Manuscripts describing research involving non-human primates must include details of animal welfare, including information about housing, feeding, and environmental enrichment, and steps taken to minimize suffering, including use of anesthesia and method of sacrifice if appropriate, in accordance with the recommendations of the Weatherall report, [The use of non-human primates in research \(PDF\)](#).

Random source animals

Manuscripts describing studies that use random source (e.g. Class B dealer-sourced in the USA), shelter, or stray animals will be subject to additional ethics consideration and may be rejected if sufficient ethical and scientific justification for the study design is lacking.

Unacceptable euthanasia methods and anesthetic agents

Manuscripts reporting use of a euthanasia method(s) classified as unacceptable by the [American Veterinary Medical Association](#) (e.g., chloral hydrate, ether, chloroform) will not be considered at *PLOS ONE* unless authors also provide, at the time of initial submission, scientific and ethical justification for use in the specific study design, as well as confirmation of approval for specific use from their Institutional Animal Care and Use Committee (IACUC) or animal research ethics committee. Manuscripts reporting use of an anesthesia method(s) that is widely prohibited or of potential concern (chloral hydrate, ether, chloroform) should include a statement of scientific and ethical justification for use in the specific study design, as well as confirmation of approval for specific use from the authors' IACUC or animal research ethics committee. These manuscripts may be subject to additional ethics considerations prior to publication.

For additional guidance on appropriate euthanasia methods, authors may also refer to:

- [Annex IV of the EU Directive 2010/EU/63 \(PDF\)](#)
- [CCAC Guidelines: on euthanasia of animals used in science \(PDF\)](#)
- [Report on the Second Newcastle Meeting on Laboratory Animal Euthanasia](#)

Humane endpoints

For studies in which death of a regulated animal (vertebrate, cephalopod) is a likely outcome or a planned experimental endpoint, *PLOS ONE* asks authors to report additional details related to the study design. This applies to research that involves, for instance, assessment of survival, toxicity, longevity, terminal disease, or high rates of incidental mortality. These studies may be subject to additional ethical considerations, and *PLOS ONE* may reject submissions if they lack sufficient reporting, appropriate justification for the study design, or adequate consideration of humane endpoints, regardless of study-specific institutional animal ethics committee approval.

Definition of a humane endpoint

A humane endpoint is an experimental endpoint at which animals are euthanized when they display early markers associated with death or poor prognosis of quality of life, or specific signs of severe suffering or distress. Humane endpoints are used as an alternative to allowing such conditions to continue or progress to death following the experimental intervention (“death as an endpoint”), or only euthanizing animals at the end of an experiment. Before a study begins, researchers define the practical observations or measurements that will be used during the study to recognize a humane endpoint, based on anticipated clinical, physiological, and behavioral signs. These may include, for instance, body temperature or weight changes, tumor size or appearance, abnormal behaviors, pathological changes, ruffled fur, reduced mobility, body posture, or expression of specific body fluid markers. Please see the [NC3Rs guidelines](#) for more information.

Authors of these studies should report all of the following information in the Methods section:

1. Describe whether humane endpoints were used for all animals involved in the study	
<i>If humane endpoints were used, report the following:</i> • The specific criteria used to determine when animals should be euthanized • Once animals reached endpoint criteria, the amount of time elapsed before euthanasia • Whether any animals died before meeting criteria for euthanasia	<i>If humane endpoints were not used, report the following:</i> • A scientific and ethical justification for the study design, including the reasons why humane endpoints could not be used, and discussion of alternatives that were considered but could not be used • Whether the institutional animal ethics committee specifically reviewed and approved the anticipated mortality in the

1. Describe whether humane endpoints were used for all animals involved in the study
study design
2. Include the following details of the study design and outcomes:
• The duration of the experiment • The numbers of animals used, euthanized, and found dead (if any); the cause of death for all animals • How frequently animal health and behavior were monitored • All animal welfare considerations taken, including efforts to minimize suffering and distress, use of analgesics or anaesthetics, or special housing conditions • Any special training in animal care or handling provided for research staff

Observational and field studies

Methods sections for submissions reporting on any type of field study must include ethics statements that specify:

- Permits and approvals obtained for the work, including the full name of the authority that approved the study; if none were required, authors should explain why
- Whether the land accessed is privately owned or protected
- Whether any protected species were sampled
- Full details of animal husbandry, experimentation, and care/welfare, where relevant

Paleontology and archaeology research

Manuscripts reporting paleontology and archaeology research must include descriptions of methods and specimens in sufficient detail to allow the work to be reproduced. Data sets supporting statistical and phylogenetic analyses should be provided, preferably in a format that allows easy re-use. [Read the policy.](#)

Specimen numbers and complete repository information, including museum name and geographic location, are required for publication. Locality information should be provided in the manuscript as legally allowable, or a statement should be included giving details of the availability of such information to qualified researchers.

If permits were required for any aspect of the work, details should be given of all permits that were obtained, including the full name of the issuing authority. This should be accompanied by the following statement:

All necessary permits were obtained for the described study, which complied with all relevant regulations.

If no permits were required, please include the following statement:

No permits were required for the described study, which complied with all relevant regulations.

Manuscripts describing paleontology and archaeology research are subject to the following policies:

- **Sharing of data and materials.** Any specimen that is erected as a new species, described, or figured must be deposited in an accessible, permanent repository (i.e., public museum or similar institution). If study conclusions depend on specimens that do not fit these criteria, the article will be rejected under *PLOS ONE*'s [data availability criterion](#).
- **Ethics.** *PLOS ONE* will not publish research on specimens that were obtained without necessary permission or were illegally exported.

Systematic reviews and meta-analyses

A systematic review paper, as defined by [The Cochrane Collaboration](#), is a review of a clearly formulated question that uses explicit, systematic methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review. These reviews differ substantially from narrative-based reviews or synthesis articles. Statistical methods (meta-analysis) may or may not be used to analyze and summarize the results of the included studies.

Reports of systematic reviews and meta-analyses must include a completed [PRISMA \(Preferred Reporting Items for Systematic Reviews and Meta-Analyses\)](#) checklist and flow diagram to accompany the main text. Blank templates are available here:

- Checklist: [PDF](#) or [Word document](#)
- Flow diagram: [PDF](#) or [Word document](#)

Authors must also state in their “Methods” section whether a protocol exists for their systematic review, and if so, provide a copy of the protocol as supporting information and provide the registry number in the abstract.

If your article is a systematic review or a meta-analysis you should:

- State this in your cover letter
- Select “Research Article” as your article type when submitting
- Include the PRISMA flow diagram as Fig 1 (required where applicable)
- Include the PRISMA checklist as supporting information

Meta-analysis of genetic association studies

Manuscripts reporting a meta-analysis of genetic association studies must report results of value to the field and should be reported according to the guidelines presented in [Systematic Reviews of Genetic Association Studies](#) by Sagoo *et al.*

On submission, authors will be asked to justify the rationale for the meta-analysis and how it contributes to the base of scientific knowledge in the light of previously published results. Authors will also be asked to complete a [checklist \(DOCX\)](#) outlining information about the justification for the study and the methodology employed. Meta-analyses that replicate published studies will be rejected if the authors do not provide adequate justification.

Personal data from third-party sources

For all studies using personal data from internet-based and other third-party sources (e.g., social media, blogs, other internet sources, mobile phone companies), data must be collected and used according to company/website Terms and Conditions, with appropriate permissions. All data sources must be acknowledged clearly in the [Materials and Methods section](#).

[Read our policy on data availability.](#)

In the Ethics Statement, authors should declare any potential risks to individuals or individual privacy, or affirm that in their assessment, the study posed no such risks. In addition, the following Ethics and Data Protection requirements must be met.

For interventional studies, which impact participants’ experiences or data, the study design must have been prospectively approved by an Ethics Committee, and informed consent is required. The Ethics Committee may waive the requirement for approval and/or consent.

For observational studies in which personal experiences and accounts are not manipulated, consultation with an Ethics or Data Protection Committee is recommended. Additional requirements apply in the following circumstances:

- If information used could threaten personal privacy or damage the reputation of individuals whose data are used, an Ethics Committee should be consulted and informed consent obtained or specifically addressed.

- If authors accessed any personal identifying information, an Ethics or Data Protection Committee should oversee data anonymization. If data were anonymized and/or aggregated before access and analysis, informed consent is generally not required.

Note that Terms of Use contracts do not qualify as informed consent, even if they address the use of personal data for research.

[See our reporting guidelines for human subjects research.](#)

Cell lines

Authors reporting research using cell lines should state when and where they obtained the cells, giving the date and the name of the researcher, cell line repository, or commercial source (company) who provided the cells, as appropriate.

Authors must also include the following information for each cell line:

For *de novo* (new) cell lines, including those given to the researchers as a gift, authors must follow our policies for [human subjects research](#) or [animal research](#), as appropriate. The ethics statement must include:

- Details of institutional review board or ethics committee approval; AND
- For human cells, confirmation of written informed consent from the donor, guardian, or next of kin

For established cell lines, the Methods section should include:

- A reference to the published article that first described the cell line; AND/OR
- The cell line repository or company the cell line was obtained from, the catalogue number, and whether the cell line was obtained directly from the repository/company or from another laboratory

Authors should check established cell lines using the [ICLAC Database of Cross-contaminated or Misidentified Cell Lines](#) to confirm they are not misidentified or contaminated. Cell line authentication is recommended – e.g., by karyotyping, isozyme analysis, or short tandem repeats (STR) analysis – and may be required during peer review or after publication.

Blots and gels

Manuscripts reporting results from blots (including Western blots) and electrophoretic gels should follow these guidelines:

- [In accordance with our policy on image manipulation](#), the image should not be adjusted in any way that could affect the scientific information displayed, e.g. by modifying the background or contrast.
- All blots and gels that support results reported in the manuscript should be provided.
- Original uncropped and unadjusted blots and gels, including molecular size markers, should be provided in either the figures or the supplementary files.
- Lanes should not be overcropped around the bands; the image should show most or all of the blot or gel. Any non-specific bands should be shown and an explanation of their nature should be given.
- The image should include all relevant controls, and controls should be run on the same blot or gel as the samples.
- A figure panel should not include composite images of bands originating from different blots or gels. If the figure shows non-adjacent bands from the same blot or gel, this should be clearly denoted by vertical black lines and the figure legend should provide details of how the figure was made.

Antibodies

Manuscripts reporting experiments using antibodies should include the following information:

- The name of each antibody, a description of whether it is monoclonal or polyclonal, and the host species.
- The commercial supplier or source laboratory.
- The catalogue or clone number and, if known, the batch number.
- The antigen(s) used to raise the antibody.
- For established antibodies, a stable public identifier from the [Antibody Registry](#).

The manuscript should also report the following experimental details:

- The final antibody concentration or dilution.
- A reference to the validation study if the antibody was previously validated. If not, provide details of how the authors validated the antibody for the applications and species used.

We encourage authors to consider adding information on new validations to a publicly available database such as [Antibodypedia](#) or [CiteAb](#).

Small and macromolecule crystal data

Manuscripts reporting new and unpublished three-dimensional structures must include sufficient supporting data and detailed descriptions of the methodologies used to allow the reproduction and validation of the structures. All novel structures must have been deposited in a community endorsed database prior to submission (please see our list of [recommended repositories](#)).

Small molecule single crystal data

Authors reporting X-Ray crystallographic structures of small organic, metal-organic, and inorganic molecules must deposit their data with the Cambridge Crystallographic Data Centre (CCDC), the Inorganic Crystal Structure Database (ICSD), or similar community databases providing a recognized validation functionality. Authors are also required to include the relevant structure reference numbers within the main text (e.g. the CCDC ID number), as well as the crystallographic information files (.cif format) as Supplementary Information, along with the checkCIF validation reports that can be obtained via the International Union of Crystallography (IUCr).

Macromolecular structures

Authors reporting novel macromolecular structures must have deposited their data prior to submission with the Worldwide Protein Data Bank (wwPDB), the Biological Magnetic Resonance Data Bank (BMRB), the Electron Microscopy Data Bank (EMDB), or other community databases providing a recognized validation functionality. Authors must include the structure reference numbers within the main text and submit as Supplementary Information the official validation reports from these databases.

Methods, software, databases, and tools

PLOS ONE will consider submissions that present new methods, software, or databases as the primary focus of the manuscript if they meet the following criteria:

Utility

The tool must be of use to the community and must present a proven advantage over existing alternatives, where applicable. Recapitulation of existing methods, software, or databases is not useful and will not be considered for publication. Combining data and/or functionalities from other sources may be acceptable, but simpler instances (i.e. presenting a subset of an already existing database) may not be considered. For software, databases, and online tools, the long-term utility should also be discussed, as relevant. This discussion may include maintenance, the potential for future growth, and the stability of the hosting, as applicable.

Validation

Submissions presenting methods, software, databases, or tools must demonstrate that the new tool achieves its intended purpose. If similar options already exist, the submitted manuscript must demonstrate that the new tool is an

improvement over existing options in some way. This requirement may be met by including a proof-of-principle experiment or analysis; if this is not possible, a discussion of the possible applications and some preliminary analysis may be sufficient.

Availability

Software should be open source, deposited in an appropriate archive, and conform to the [Open Source Definition](#). Databases must be open-access and hosted somewhere publicly accessible, and any software used to generate a database should also be open source. If relevant, databases should be open for appropriate deposition of additional data. Dependency on commercial software such as Mathematica and MATLAB does not preclude a paper from consideration, although complete open source solutions are preferred. Authors should provide a direct link to the deposited software or the database hosting site from within the paper.

Software submissions

Manuscripts describing software should provide full details of the algorithms designed. Describe any dependencies on commercial products or operating system. Include details of the supplied test data and explain how to install and run the software. A brief description of enhancements made in the major releases of the software may also be given. Authors should provide a direct link to the deposited software from within the paper.

Database submissions

For descriptions of databases, provide details about how the data were curated, as well as plans for long-term database maintenance, growth, and stability. Authors should provide a direct link to the database hosting site from within the paper.

New taxon names

Zoological names

When publishing papers that describe a new zoological taxon name, PLOS aims to comply with the requirements of the [International Commission on Zoological Nomenclature \(ICZN\)](#). Effective 1 January 2012, the ICZN considers an online-only publication to be legitimate if it meets the criteria of archiving and is registered in ZooBank, the ICZN's official registry.

For proper registration of a new zoological taxon, we require two specific statements to be included in your manuscript.

In the **Results** section, the globally unique identifier (GUID), currently in the form of a Life Science Identifier (LSID), should be listed under the new species name, for example:

Anochetus boltoni Fisher **sp. nov.** urn:lsid:zoobank.org:act:B6C072CF-1CA6-40C7-8396-534E91EF7FBB

You will need to contact [ZooBank](#) to obtain a GUID (LSID). Please do this as early as possible to avoid delay of publication upon acceptance of your manuscript. It is your responsibility to provide us with this information so we can include it in the final published paper.

Please also insert the following text into the **Methods** section, in a sub-section to be called “Nomenclatural Acts”:

The electronic edition of this article conforms to the requirements of the amended International Code of Zoological Nomenclature, and hence the new names contained herein are available under that Code from the electronic edition of this article. This published work and the nomenclatural acts it contains have been registered in ZooBank, the online registration system for the ICZN. The ZooBank LSIDs (Life Science Identifiers) can be resolved and the associated information viewed through any standard web browser by appending the LSID to the prefix “<http://zoobank.org/>”. The LSID for this publication is: urn:lsid:zoobank.org:pub: XXXXXXXX. The electronic edition of this work was published in a journal with an ISSN, and has been archived and is available from the following digital repositories: PubMed Central, LOCKSS [author to insert any additional repositories].

All PLOS articles are deposited in [PubMed Central](#) and [LOCKSS](#). If your institute, or those of your co-authors, has its own repository, we recommend that you also deposit the published online article there and include the name in your article.

Botanical names

When publishing papers that describe a new botanical taxon, PLOS aims to comply with the requirements of the International Code of Nomenclature for algae, fungi, and plants (ICN). The following guidelines for publication in an online-only journal have been agreed such that any scientific botanical name published by us is considered effectively published under the rules of the Code. Please note that these guidelines differ from those for zoological nomenclature, and apply only to seed plants, ferns, and lycophytes.

Effective January 2012, the description or diagnosis of a new taxon can be in either Latin or English. This does not affect the requirements for scientific names, which are still to be Latin.

Also effective January 2012, the electronic PDF represents a published work according to the ICN for algae, fungi, and plants. Therefore the new names contained in the electronic publication of PLOS article are effectively published under that Code from the electronic edition alone, so there is no longer any need to provide printed copies.

Additional information describing recent changes to the Code can be found [here](#).

For proper registration of the new taxon, we require two specific statements to be included in your manuscript.

In the **Results** section, the globally unique identifier (GUID), currently in the form of a Life Science Identifier (LSID), should be listed under the new species name, for example:

Solanum aspersum S.Knapp, sp. nov. [urn:lsid:ipni.org:names:77103633-1] Type: Colombia. Putumayo: vertiente oriental de la Cordillera, entre Sachamates y San Francisco de Sibundoy, 1600-1750 m, 30 Dec 1940, J. Cuatrecasas 11471 (holotype, COL; isotypes, F [F-1335119], US [US-1799731]).

Journal staff will contact IPNI to obtain the GUID (LSID) after your manuscript is accepted for publication, and this information will then be added to the manuscript during the production phase

In the **Methods** section, include a sub-section called “Nomenclature” using the following wording:

The electronic version of this article in Portable Document Format (PDF) in a work with an ISSN or ISBN will represent a published work according to the International Code of Nomenclature for algae, fungi, and plants, and hence the new names contained in the electronic publication of a PLOS article are effectively published under that Code from the electronic edition alone, so there is no longer any need to provide printed copies.

In addition, new names contained in this work have been submitted to IPNI, from where they will be made available to the Global Names Index. The IPNI LSIDs can be resolved and the associated information viewed through any standard web browser by appending the LSID contained in this publication to the prefix <http://ipni.org/>. The online version of this work is archived and available from the following digital repositories: [INSERT NAMES OF DIGITAL REPOSITORIES WHERE ACCEPTED MANUSCRIPT WILL BE SUBMITTED (PubMed Central, LOCKSS etc)].

All PLOS articles are deposited in [PubMed Central](#) and [LOCKSS](#). If your institute, or those of your co-authors, has its own repository, we recommend that you also deposit the published online article there and include the name in your article.

Fungal names

When publishing papers that describe a new botanical taxon, PLOS aims to comply with the requirements of the International Code of Nomenclature for algae, fungi, and plants (ICN). The following guidelines for publication in an online-only journal have been agreed such that any scientific botanical name published by us is considered effectively published under the rules of the Code. Please note that these guidelines differ from those for zoological nomenclature.

Effective January 2012, the description or diagnosis of a new taxon can be in either Latin or English. This does not affect the requirements for scientific names, which are still to be Latin.

Also effective January 2012, the electronic PDF represents a published work according to the ICN for algae, fungi, and plants. Therefore the new names contained in the electronic publication of PLOS article are effectively published under that Code from the electronic edition alone, so there is no longer any need to provide printed copies.

Additional information describing recent changes to the Code can be found [here](#).

For proper registration of the new taxon, we require two specific statements to be included in your manuscript.

In the **Results** section, the globally unique identifier (GUID), currently in the form of a Life Science Identifier (LSID), should be listed under the new species name, for example:

Hymenogaster huthii. Stielow et al. 2010, sp. nov. [urn:lsid:indexfungorum.org:names:518624]

You will need to contact either [Mycobank](#) or [Index Fungorum](#) to obtain the GUID (LSID). Please do this as early as possible to avoid delay of publication upon acceptance of your manuscript. It is your responsibility to provide us with this information so we can include it in the final published paper. Effective January 2013, all papers describing new fungal species must reference the identifier issued by a recognized repository in the protologue in order to be considered effectively published.

In the **Methods** section, include a sub-section called “Nomenclature” using the following wording (this example is for taxon names submitted to MycoBank; please substitute appropriately if you have submitted to Index Fungorum):

The electronic version of this article in Portable Document Format (PDF) in a work with an ISSN or ISBN will represent a published work according to the International Code of Nomenclature for algae, fungi, and plants, and hence the new names contained in the electronic publication of a PLOS article are effectively published under that Code from the electronic edition alone, so there is no longer any need to provide printed copies.

In addition, new names contained in this work have been submitted to MycoBank from where they will be made available to the Global Names Index. The unique MycoBank number can be resolved and the associated information viewed through any standard web browser by appending the MycoBank number contained in this publication to the prefix <http://www.mycobank.org/MB/>. The online version of this work is archived and available from the following digital repositories: [INSERT NAMES OF DIGITAL REPOSITORIES WHERE ACCEPTED MANUSCRIPT WILL BE SUBMITTED (PubMed Central, LOCKSS etc)].

All PLOS articles are deposited in [PubMed Central](#) and [LOCKSS](#). If your institute, or those of your co-authors, has its own repository, we recommend that you also deposit the published online article there and include the name in your article.

Qualitative research

Qualitative research studies use non-quantitative methods to address a defined research question that may not be accessible by quantitative methods, such as people's interpretations, experiences, and perspectives. The analysis methods are explicit, systematic, and reproducible, but the results do not involve numerical values or use statistics. Examples of qualitative data sources include, but are not limited to, interviews, text documents, audio/video recordings, and free-form answers to questionnaires and surveys.

Qualitative research studies should be reported in accordance to the [Consolidated criteria for reporting qualitative research \(COREQ\) checklist](#). Further reporting guidelines can be found in the Equator Network's [Guidelines for reporting qualitative research](#).

4.2 Scientific Reports

SCIENTIFIC REPORTS

a nature research journal

Aims & Scope

Scientific Reports is an online, open access journal from the publishers of *Nature*. We publish original scientifically valid and technically sound primary research from all areas of the natural and clinical sciences.

By publishing with us, your research will be made accessible to the research community, and get the coverage and attention it deserves. Open access and continuous online publication means your work will be published swiftly, ready to be accessed by anyone, anywhere, at any time.

Peer review

Each manuscript is handled by an Editorial Board Member, and most submissions are peer reviewed by one or more referees in addition to the Editorial Board Member. The Editorial Board — composed of practising scientists from all fields within the journal's scope — manages the peer review process, and decides whether a paper should be accepted for publication.

The aim of the peer review process is to establish the technical soundness of a submission, and uses a template approach. Referees and Editorial Board Members will determine whether a paper is scientifically valid, rather than making judgements on significance or whether the submission represents a conceptual advance.

Submission guidelines

General information for preparing manuscripts

Format of articles

Scientific Reports publishes original research in one format, Article. In most cases we do not impose strict limits on word count or page number. We do, however, strongly encourage authors to write concisely and to adhere to the guidelines below.

Articles should ideally be no more than 11 typeset pages in length. As a guide, the main text (not including Abstract, Methods, References and figure legends) should be no more than 4,500 words. The maximum Article title length is 20 words. The Abstract — which must be no more than 200 words long and contain no references — should serve both as a general introduction to the topic and as a brief, non-technical summary of the main results and their implications.

For the main body of the text, there are no explicit requirements for section organization. According to the authors' preference, the text may be organized as best suits the research. As a guideline and in the majority of cases, however, we recommend that you structure your manuscript as follows:

- Introduction
- Results (with subheadings)
- Discussion (without subheadings)
- Methods

A specific order for the main body of the text is not compulsory and, in some cases, it may be appropriate to combine sections. Figure legends are limited to 350 words. As a guideline references should be limited to 60 (this is not strictly enforced). Footnotes should not be used.

We suggest that Articles contain no more than 8 display items ([figures](#) and/ or [tables](#)). In addition, a limited number of uncaptioned molecular structure graphics and numbered mathematical equations may be included if necessary. To enable typesetting of papers, the number of display items should be commensurate with the word length — we suggest that for Articles with less than 2,000 words, no more than 4 figures/ tables should be included. Please note that schemes are not used and should be presented as figures.

Authors must provide a competing financial interests statement within the manuscript file.

Submissions should include a cover letter, a manuscript text file, individual figure files and optional supplementary information files. For first submissions (i.e. not revised manuscripts), authors may incorporate the manuscript text and figures into a single file up to 3 MB in size; the figures may be inserted in the text at the appropriate positions, or grouped at the end. Supplementary information should be combined and supplied as a single separate file, preferably in PDF format.

ONLY the following file types can be uploaded for Article text:

- txt, doc, docx, tex

A [submission template](#) is available in the [Overleaf](#) template gallery to help you prepare a LaTeX manuscript within the *Scientific Reports* formatting criteria.

Scientific Reports is read by scientists from diverse backgrounds. In addition, many are not native English speakers. Authors should, therefore, give careful thought to how their findings may be communicated clearly. Although a shared basic knowledge of science may be assumed, please bear in mind that the language and concepts that are standard in one field may be unfamiliar to non-specialists. Thus, technical jargon should be avoided and clearly explained where its use is unavoidable.

Abbreviations, particularly those that are not standard, should also be kept to a minimum. Where unavoidable, abbreviations should be defined in the text or legends at their first occurrence, and abbreviations should be used thereafter. The background, rationale and main conclusions of the study should be clearly explained. Titles and abstracts in particular should be written in language that will be readily intelligible to any scientist. We strongly recommend that authors ask a colleague with different expertise to review the manuscript before submission, in order to identify concepts and terminology that may present difficulties to non-specialist readers.

The format requirements of *Scientific Reports* are described below.

Scientific Reports uses UK English spelling.

Cover letter

Authors should provide a cover letter that includes the affiliation and contact information for the corresponding author. Authors should briefly explain why the work is considered appropriate for *Scientific Reports*. Authors are asked to suggest the names and contact information for scientific reviewers and they may request the exclusion of certain referees. Please ensure that your cover letter also includes suggestions for [Editorial Board Members](#) who would be able to handle your submission. Finally, authors should indicate whether they have had any prior discussions with a *Scientific Reports* Editorial Board Member about the work described in the manuscript.

Format of manuscripts

In most cases we do not impose strict limits on word counts and page numbers, but we encourage authors to write concisely and suggest authors adhere to the guidelines below. For a definitive list of which limits are mandatory please visit the [submission checklist page](#).

Articles should be no more than 11 typeset pages in length. As a guide, the main text (not including Abstract, Methods, References and figure legends) should be no more than 4,500 words. The maximum title length is 20 words. The Abstract (without heading) - which must be no more than 200 words long and contain no references - should serve both as a general introduction to the topic and as a brief, non-technical summary of the main results and their implications.

The manuscript text file should include the following parts, in order: a title page with author affiliations and contact information (the corresponding author should be identified with an asterisk). The main text of an Article can be organised in different ways and according to the authors' preferences, it may be appropriate to combine sections.

As a guideline, we recommend that sections include an Introduction of referenced text that expands on the background of the work. Some overlap with the Abstract is acceptable. This may then be followed by sections headed Results (with subheadings), Discussion (without subheadings) and Methods.

The main body of text must be followed by References, Acknowledgements (optional), Author Contributions (names must be given as initials), Additional Information (including a Competing Financial Interests Statement), Figure Legends (these are limited to 350 words per figure) and Tables (maximum size of one page). Footnotes are not used.

For first submissions (i.e. not revised manuscripts), authors may choose to incorporate the manuscript text and figures into a single file up to 3 MB in size - the figures may be inserted within the text at the appropriate positions, or grouped at the end. Supplementary Information should be combined and supplied as a separate file, preferably in PDF format. The first page of the Supplementary Information file should include the title of the manuscript and the author list.

Authors who do not incorporate the manuscript text and figures into a single file should adhere to the following: all textual content should be provided in a single file, prepared using either Microsoft Word or LaTeX; figures should be provided as individual files.

The manuscript file should be formatted as single-column text without justification. Pages should be numbered using an Arabic numeral in the footer of each page. Standard fonts are recommended and the 'symbols' font should be used for representing Greek characters.

TeX/ LaTeX - Authors submitting LaTeX files may use the standard 'article' document class (or similar) or may use the wlscrep.cls file and [template provided by Overleaf](#). Non-standard fonts should be avoided; please use the default Computer Modern fonts. For the inclusion of graphics, we recommend graphicx.sty. Please use numerical references only for citations. Our system cannot accept .bib files. If references are prepared using BibTeX (which is optional), please include the .bbl file with your submission (as a 'related manuscript file') in order for it to be processed correctly; this file is included automatically in the zip file generated by Overleaf for submissions. Please see this [help article on Overleaf](#) for more details. Alternatively ensure that the references (source code) are included within the manuscript file itself. As a final precaution, authors should ensure that the complete .tex file compiles successfully on their own system with no errors or warnings, before submission.

Manuscripts published in *Scientific Reports* are not subject to in-depth copy editing as part of the production process. Authors are responsible for procuring copy editing or language editing services for their manuscripts, either before submission, or at the revision stage, should they feel it would benefit their manuscript. Such services include those provided by our affiliates [Nature Research](#)

[Editing Service](#) and [American Journal Experts](#). *Scientific Reports* authors are entitled to a 10% discount on their first submission to either of these services. To claim 10% off English editing from Nature Research Editing Service, click [here](#). To claim 10% off American Journal Experts, click [here](#). Please note that the use of an editing service is at the author's own expense, and in no way implies that the article will be selected for peer-review or accepted for publication.

Methods

Where appropriate, we recommend that authors limit their Methods section to 1,500 words. Authors must ensure that their Methods section includes adequate experimental and characterization data necessary for others in the field to reproduce their work. Descriptions of standard protocols and experimental procedures should be given. Commercial suppliers of reagents or instrumentation should be identified only when the source is critical to the outcome of the experiments. Sources for kits should be identified. Experimental protocols that describe the synthesis of new compounds should be included. The systematic name of the compound and its bold Arabic numeral are used as the heading for the experimental protocol. Thereafter, the compound is represented by its assigned bold numeral. Authors should describe the experimental protocol in detail, referring to amounts of reagents in parentheses, when possible (eg 1.03 g, 0.100 mmol). Standard abbreviations for reagents and solvents are encouraged. Safety hazards posed by reagents or protocols should be identified clearly. Isolated mass and percent yields should be reported at the end of each protocol.

References

References will not be copy edited by *Scientific Reports*. References will be linked electronically to external databases where possible, making correct formatting of the references essential.

References should be numbered sequentially, first throughout the text, then in tables, followed by figures; that is, references that only appear in tables or figures should be last in the reference list. Only one publication is given for each number. Only papers or datasets that have been published or accepted by a named publication, recognized preprint server or data repository should be in the numbered list; preprints of accepted papers in the reference list should be submitted with the manuscript. Published conference abstracts and numbered patents may be included in the reference list. Grant details and acknowledgements are not permitted as numbered references. Footnotes are not used.

BibTeX (.bib) bibliography files cannot be accepted. LaTeX submission must either contain all references within the manuscript .tex file itself, or (for authors using the Overleaf template) can include the .bbl file generated during the compilation process as a 'related manuscript file' (see the "Format of manuscripts" section for more details).

Scientific Reports uses standard *Nature* referencing style. All authors should be included in reference lists unless there are six or more, in which case only the first author should be given, followed by 'et al.'. Authors should be listed last name first, followed by a comma and initials (followed by full stops) of given names. Article and dataset titles should be in Roman text, only the first word of the title should have an initial capital and the title should be written exactly as it appears in the work cited, ending with a full stop. Book titles should be given in italics and all words in the title should have initial capitals. Journal and data repository names are italicized and abbreviated (with full stops) according to common usage. Volume numbers and the subsequent comma appear in bold. The full page range should be given (or article number), where appropriate.

Published papers:

Printed journals

Schott, D. H., Collins, R. N. & Bretscher, A. Secretory vesicle transport velocity in living cells depends on the myosin V lever arm length. *J. Cell Biol.* **156**, 35-39 (2002).

Online only

Bellin, D. L. *et al.* Electrochemical camera chip for simultaneous imaging of multiple metabolites in biofilms. *Nat. Commun.* **7**, 10535; [10.1038/ncomms10535](https://doi.org/10.1038/ncomms10535) (2016).

For papers with more than five authors include only the first author's name followed by 'et al.'.

Books:

Smith, J. Syntax of referencing in *How to reference books* (ed. Smith, S.) 180-181 (Macmillan, 2013).

Online material:

Babichev, S. A., Ries, J. & Lvovsky, A. I. Quantum scissors: teleportation of single-mode optical states by means of a nonlocal single photon. Preprint at <https://arxiv.org/abs/quant-ph/0208066> (2002).

Manaster, J. Sloth squeak. *Scientific American Blog Network* <http://blogs.scientificamerican.com/psi-vid/2014/04/09/sloth-squeak> (2014).

Hao, Z., AghaKouchak, A., Nakhjiri, N. & Farahmand, A. Global integrated drought monitoring and prediction system (GIDMaPS) data sets. *figshare* <https://doi.org/10.6084/m9.figshare.853801> (2014).

Acknowledgements

Acknowledgements should be brief, and should not include thanks to anonymous referees and editors, or effusive comments. Grant or contribution numbers may be acknowledged. Assistance from medical writers, proof-readers and editors should also be acknowledged here.

Author contributions

Scientific Reports requires an Author Contribution Statement as described in the [Author responsibilities](#) section of our [Editorial and Publishing Policies](#).

Competing financial interests

A [competing financial interests statement](#) is required for all accepted papers published in *Scientific Reports*. If there is no conflict of interest, a statement declaring this will still be included in the paper.

Data availability

Scientific Reports requires a Data Availability Statement to be included in the Methods section of submitted manuscripts (see '[Availability of materials and data](#)' section for more information).

Supplementary Information

Any Supplementary Information should be submitted with the manuscript and will be sent to referees during peer review. It is published online with accepted manuscripts. We request that authors avoid "data not shown" statements and instead make their data available via deposition in a public repository (see '[Availability of materials and data](#)' for more information). Any data necessary to evaluation of the claims of the paper that are not available via a public depository should be provided as Supplementary Information. Supplementary Information is not edited, typeset or proofed, so authors should ensure that it is clearly and succinctly presented at initial submission, and that the style and terminology conform to the rest of the paper. Authors should include the title of the manuscript and full author list on the first page.

The guidelines below detail the creation, citation and submission of Supplementary Information - publication may be delayed if these are not followed correctly. Please note that modification of

Supplementary Information after the paper is published requires a formal correction, so authors are encouraged to check their Supplementary Information carefully before submitting the final version.

1. Multiple pieces of Supplementary Information can be combined and supplied as a single file, or supplied separately (e.g. supplementary videos, spreadsheets [.csv or .xlsx] or data files).
2. Designate each item as Supplementary Table, Figure, Video, Audio, Note, Data, Discussion, Equations or Methods, as appropriate. Number Supplementary Tables and Figures as, for example, "Supplementary Table S1". This numbering should be separate from that used in tables and figures appearing in the main article. Supplementary Note or Methods should not be numbered; titles for these are optional.
3. Refer to each piece of supplementary material at the appropriate point(s) in the main article. Be sure to include the word "Supplementary" each time one is mentioned. Please do not refer to individual panels of supplementary figures.
4. Use the following examples as a guide (note: abbreviate "Figure" as "Fig." when in the middle of a sentence): "Table 1 provides a selected subset of the most active compounds. The entire list of 96 compounds can be found as Supplementary Table S1 online." "The biosynthetic pathway of L-ascorbic acid in animals involves intermediates of the D-glucuronic acid pathway (see Supplementary Fig. S2 online). Figure 2 shows..."
5. Remember to include a brief title and legend (incorporated into the file to appear near the image) as part of every figure submitted, and a title as part of every table.
6. File sizes should be as small as possible, with a maximum size of 50 MB, so that they can be downloaded quickly.

Further queries about submission and preparation of Supplementary Information should be directed to email: scirep.admin@nature.com.

Figure legends

Figure legends begin with a brief title sentence for the whole figure and continue with a short description of what is shown in each panel in sequence and the symbols used; methodological details should be minimised as much as possible. Each legend must total no more than 350 words. Text for figure legends should be provided in numerical order after the references.

Tables

Please submit tables in your main article document in an editable format (Word or TeX/ LaTeX, as appropriate), and not as images. Tables that include statistical analysis of data should describe their standards of error analysis and ranges in a table legend.

Equations

Equations and mathematical expressions should be provided in the main text of the paper. Equations that are referred to in the text are identified by parenthetical numbers, such as (1), and are referred to in the manuscript as "equation (1)".

If your manuscript is or will be in .docx format and contains equations, you must follow the instructions below to make sure that your equations are editable when the file enters production.

If you have not yet composed your article, you can ensure that the equations in your .docx file remain editable in .doc by enabling "Compatibility Mode" before you begin. To do this, open a new document and save as Word 97-2003 (*.doc). Several features of Word 2007/ 10 will now be inactive, including the built-in equation editing tool. You can insert equations in one of the two ways listed below.

If you have already composed your article as .docx and used its built-in equation editing tool, your equations will become images when the file is saved down to .doc. To resolve this problem, re-key your equations in one of the two following ways.

1. Use MathType to create the equation. MathType is the recommended method for creating equations.
2. Go to Insert > Object > Microsoft Equation 3.0 and create the equation.

If, when saving your final document, you see a message saying "Equations will be converted to images", your equations are no longer editable and we will not be able to accept your file.

General figure guidelines

Authors are responsible for obtaining permission to publish any figures or illustrations that are protected by copyright, including figures published elsewhere and pictures taken by professional photographers. The journal cannot publish images downloaded from the internet without appropriate permission.

Figures should be numbered separately with Arabic numerals in the order of occurrence in the text of the manuscript. When appropriate, figures should include error bars. A description of the statistical treatment of error analysis should be included in the figure legend. Please note that schemes are not used; sequences of chemical reactions or experimental procedures should be submitted as figures, with appropriate captions. A limited number of uncaptioned graphics depicting chemical structures - each labelled with their name, by a defined abbreviation, or by the bold Arabic numeral - may be included in a manuscript.

Figure lettering should be in a clear, sans-serif typeface (for example, Helvetica); the same typeface in the same font size should be used for all figures in a paper. Use 'symbols' font for Greek letters. All display items should be on a white background, and should avoid excessive boxing, unnecessary colour, spurious decorative effects (such as three-dimensional 'skyscraper' histograms) and highly pixelated computer drawings. The vertical axis of histograms should not be truncated to exaggerate small differences. Labelling must be of sufficient size and contrast to be readable, even after appropriate reduction. The thinnest lines in the final figure should be no smaller than one point wide. Authors will see a proof that will include figures.

Figures divided into parts should be labelled with a lower-case bold a, b, and so on, in the same type size as used elsewhere in the figure. Lettering in figures should be in lower-case type, with only the first letter of each label capitalized. Units should have a single space between the number and the unit, and follow SI nomenclature (for example, ms rather than msec) or the nomenclature common to a particular field. Thousands should be separated by commas (1,000). Unusual units or abbreviations should be spelled out in full or defined in the legend. Scale bars should be used rather than magnification factors, with the length of the bar defined on the bar itself rather than in the legend. In legends, please use visual cues rather than verbal explanations such as "open red triangles".

Unnecessary figures should be avoided: data presented in small tables or histograms, for instance, can generally be stated briefly in the text instead. Figures should not contain more than one panel unless the parts are logically connected; each panel of a multipart figure should be sized so that the whole figure can be reduced by the same amount and reproduced at the smallest size at which essential details are visible.

Figures for peer review

At the initial submission stage authors may choose to upload separate figure files or to incorporate figures into the main article file, ensuring that any inserted figures are of sufficient quality to be clearly legible.

When submitting a revised manuscript all figures must be uploaded as separate figure files ensuring that the image quality and formatting conforms to the specifications below.

Figures for publication

Each complete figure must be supplied as a separate file upload. Multi-part/ panel figures must be prepared and arranged as a single image file (including all sub-parts; a, b, c, etc.). Please do not upload each panel individually.

Please read the [digital images integrity and standards](#) section of our [Editorial and Publishing Policies](#). When possible, we prefer to use original digital figures to ensure the highest-quality reproduction in the journal. For optimal results, prepare figures to fit A4 page-width. When creating and submitting digital files, please follow the guidelines below. Failure to do so, or to adhere to the following guidelines, can significantly delay publication of your work.

Authors are responsible for obtaining permission to publish any figures or illustrations that are protected by copyright, including figures published elsewhere and pictures taken by professional photographers. The journal cannot publish images downloaded from the internet without appropriate permission.

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Line art, graphs, charts and schematics

For optimal results, all line art, graphs, charts and schematics should be supplied in vector format, such as EPS or AI, and should be saved or exported as such directly from the application in which they were made. Please ensure that data points and axis labels are clearly legible.

2. Photographic and bitmap images

All photographic and bitmap images should be supplied in a bitmap image format such as tiff, jpg, or psd. If saving tiff files, please ensure that the compression option is selected to avoid very large file sizes.

Please do not supply Word or Powerpoint files with placed images. Images can be supplied as RGB or CMYK (note: we will not convert image colour modes).

Figures that do not meet these standards will not reproduce well and may delay publication until we receive high-resolution images.

3. Chemical structures

Chemical structures should be produced using ChemDraw or a similar program. All chemical compounds must be assigned a bold, Arabic numeral in the order in which the compounds are presented in the manuscript text. Structures should then be exported into a 300 dpi RGB tiff file before being submitted.

4. Stereo images

Stereo diagrams should be presented for divergent 'wall-eyed' viewing, with the two panels separated by 5.5 cm. In the final accepted version of the manuscript, the stereo images should be submitted at their final page size.

Statistical guidelines

Every article that contains statistical testing should state the name of the statistical test, the n value for each statistical analysis, the comparisons of interest, a justification for the use of that test (including, for example, a discussion of the normality of the data when the test is appropriate only for normal data), the alpha level for all tests, whether the tests were one-tailed or two-tailed, and the actual P value for each test (not merely "significant" or " $P < 0.05$ "). It should be clear what statistical test was

used to generate every P value. Use of the word "significant" should always be accompanied by a P value; otherwise, use "substantial," "considerable," etc.

Data sets should be summarized with descriptive statistics, which should include the n value for each data set, a clearly labelled measure of centre (such as the mean or the median), and a clearly labelled measure of variability (such as standard deviation or range). Ranges are more appropriate than standard deviations or standard errors for small data sets. Graphs should include clearly labelled error bars. Authors must state whether a number that follows the $\pm$ sign is a standard error (s.e.m.) or a standard deviation (s.d.).

Authors must justify the use of a particular test and explain whether their data conform to the assumptions of the tests. Three errors are particularly common:

- **Multiple comparisons:** When making multiple statistical comparisons on a single data set, authors should explain how they adjusted the alpha level to avoid an inflated Type I error rate, or they should select statistical tests appropriate for multiple groups (such as ANOVA rather than a series of t-tests).
- **Normal distribution:** Many statistical tests require that the data be approximately normally distributed; when using these tests, authors should explain how they tested their data for normality. If the data do not meet the assumptions of the test, then a non-parametric alternative should be used instead.
- **Small sample size:** When the sample size is small (less than about 10), authors should use tests appropriate to small samples or justify their use of large-sample tests.

There is a [checklist](#) available to help authors minimize the chance of statistical errors.

Chemical and biological nomenclature and abbreviations

Molecular structures are identified by bold, Arabic numerals assigned in order of presentation in the text. Once identified in the main text or a figure, compounds may be referred to by their name, by a defined abbreviation, or by the bold Arabic numeral (as long as the compound is referred to consistently as one of these three).

When possible, authors should refer to chemical compounds and biomolecules using systematic nomenclature, preferably using [IUPAC](#). Standard chemical and biological abbreviations should be used. Unconventional or specialist abbreviations should be defined at their first occurrence in the text.

Gene nomenclature

Authors should use approved nomenclature for gene symbols, and use symbols rather than italicized full names (for example Ttn, not titin). Please consult the appropriate nomenclature databases for correct gene names and symbols. A useful resource is LocusLink.

Approved human gene symbols are provided by HUGO Gene Nomenclature Committee (HGNC), e-mail: hgnc@genenames.org; see also www.genenames.org. Approved mouse symbols are provided by The Jackson Laboratory, e-mail: nomen@informatics.jax.org; see also www.informatics.jax.org/mgihome/nomen.

For proposed gene names that are not already approved, please submit the gene symbols to the appropriate nomenclature committees as soon as possible, as these must be deposited and approved before publication of an article.

Avoid listing multiple names of genes (or proteins) separated by a slash, as in 'Oct4/ Pou5f1', as this is ambiguous (it could mean a ratio, a complex, alternative names or different subunits). Use one name throughout and include the other at first mention: 'Oct4 (also known as Pou5f1)'.

Characterization of chemical and biomolecular materials

Scientific Reports is committed to publishing technically sound research. Manuscripts submitted to the journal will be held to rigorous standards with respect to experimental methods and characterization of new compounds. Authors must provide adequate data to support their assignment of identity and purity for each new compound described in the manuscript. Authors should provide a statement confirming the source, identity and purity of known compounds that are central to the scientific study, even if they are purchased or resynthesized using published methods.

1. Chemical identity

Chemical identity for organic and organometallic compounds should be established through spectroscopic analysis. Standard peak listings (see formatting guidelines below) for ^1H NMR and proton-decoupled ^{13}C NMR should be provided for all new compounds. Other NMR data should be reported (^{31}P NMR, ^{19}F NMR, etc.) when appropriate. For new materials, authors should also provide mass spectral data to support molecular weight identity. High-resolution mass spectral (HRMS) data are preferred. UV or IR spectral data may be reported for the identification of characteristic functional groups, when appropriate. Melting-point ranges should be provided for crystalline materials. Specific rotations may be reported for chiral compounds. Authors should provide references, rather than detailed procedures, for known compounds, unless their protocols represent a departure from or improvement on published methods.

2. Combinational compound libraries

Authors describing the preparation of combinatorial libraries should include standard characterization data for a diverse panel of library components.

3. Biomolecular identity

For new biopolymeric materials (oligosaccharides, peptides, nucleic acids, etc.), direct structural analysis by NMR spectroscopic methods may not be possible. In these cases, authors must provide evidence of identity based on sequence (when appropriate) and mass spectral characterization.

4. Biological constructs

Authors should provide sequencing or functional data that validates the identity of their biological constructs (plasmids, fusion proteins, site-directed mutants, etc.) either in the manuscript text or the Methods section, as appropriate.

5. Sample purity

Evidence of sample purity is requested for each new compound. Methods for purity analysis depend on the compound class. For most organic and organometallic compounds, purity may be demonstrated by high-field ^1H NMR or ^{13}C NMR data, although elemental analysis ($\pm 0.4\%$) is encouraged for small molecules. Quantitative analytical methods including chromatographic (GC, HPLC, etc.) or electrophoretic analyses may be used to demonstrate purity for small molecules and polymeric materials.

6. Spectral data

Detailed spectral data for new compounds should be provided in list form (see below) in the Methods section. Figures containing spectra generally will not be published as a manuscript figure unless the data are directly relevant to the central conclusions of the paper. Authors are encouraged to include high-quality images of spectral data for key compounds in the Supplementary Information. Specific NMR assignments should be listed after integration values only if they were unambiguously determined by multidimensional NMR or decoupling experiments. Authors should provide information about how assignments were made in a general Methods section.

Example format for compound characterization data. mp: 100-102 °C (lit.^{ref} 99-101 °C); TLC (CHCl_3 :MeOH, 98:2 v/v): R_f = 0.23; $[\alpha]_D$ = -21.5 (0.1 M in n-hexane); ^1H NMR (400 MHz, CDCl_3): δ 9.30

(s, 1H), 7.55-7.41 (m, 6H), 5.61 (d, J = 5.5 Hz, 1H), 5.40 (d, J = 5.5 Hz, 1H), 4.93 (m, 1H), 4.20 (q, J = 8.5 Hz, 2H), 2.11 (s, 3H), 1.25 (t, J = 8.5 Hz, 3H); ^{13}C NMR (125 MHz, CDCl_3): δ 165.4, 165.0, 140.5, 138.7, 131.5, 129.2, 118.6, 84.2, 75.8, 66.7, 37.9, 20.1; IR (Nujol): 1765 cm^{-1} ; UV/ Vis: λ_{max} 267 nm; HRMS (m/z): $[\text{M}]^+$ calcd. for $\text{C}_{20}\text{H}_{15}\text{Cl}_2\text{NO}_5$, 420.0406; found, 420.0412; analysis (calcd., found for $\text{C}_{20}\text{H}_{15}\text{Cl}_2\text{NO}_5$): C (57.16, 57.22), H (3.60, 3.61), Cl (16.87, 16.88), N (3.33, 3.33), O (19.04, 19.09).

7. Crystallographic data for small molecules

Manuscripts reporting new three-dimensional structures of small molecules from crystallographic analysis should include a .cif file and a structural figure with probability ellipsoids for publication as Supplementary Information. These must have been checked using the IUCR's [CheckCIF](#) routine, and a PDF copy of the output must be included with the submission, together with a justification for any alerts reported. Crystallographic data for small molecules should be submitted to the [Cambridge Structural Database](#) and the deposition number referenced appropriately in the manuscript. Full access must be provided on publication.

8. Macromolecular structural data

Manuscripts reporting new structures should contain a table summarizing structural and refinement statistics. Templates are available for such tables describing [NMR](#) and [X-ray crystallography](#) data. To facilitate assessment of the quality of the structural data, a stereo image of a portion of the electron density map (for crystallography papers) or of the superimposed lowest energy structures (≥ 10 ; for NMR papers) should be provided with the submitted manuscript. If the reported structure represents a novel overall fold, a stereo image of the entire structure (as a backbone trace) should also be provided.

5 Permission and contributions of co-authors

5.1 PLOS ONE (1)

None of the co-authors of the paper titled “**Contribution towards a metabolite profile of the detoxification of benzoic acid through glycine conjugation: an intervention study**” had any objection to the use of the paper, as presented in Chapter 4 (section 4.4.2), as part of the thesis submitted for the degree *Philosophiae Doctor*.

This investigation required a multidisciplinary approach and the coordinated inputs of all authors were essential to produce the outcome of the study. The contributions of the co-authors for this publication were as follows: Prof Carolus Reinecke and Prof Johan Westerhuis defined the aim of the study, developed the methodological design and supervised the planning and execution of the investigation. Prof Carolus Reinecke also obtained research funds for the project, ethical approval from the Health Sciences Ethical Committee of North-West University, and written informed consent from the 24 participants of the cross-over study. Ms Cindy Irwin conducted all experimental analyses and data collection. Dr Shayne Mason gave guidance and assessment on ^1H -NMR data generation and spectral analyses. Ms Cindy Irwin and Dr Shayne Mason were responsible for the data curation, peak annotation and production of the metadata. Ms Mari van Reenen and Prof Johan Westerhuis conceptualized the complementary techniques required to analyse the data and Ms Mari van Reenen applied selected computer programmes, implemented computer codes and supporting algorithms, and performed the statistical and computational analyses of the original and curated data. Prof Lodewyk Mienie participated throughout the investigation as expert on metabolism and in modelling of the proposed metabolic outcome of the study. All authors gave input according to personal expertise in the development of the manuscript, gave final approval of the version as submitted and agreed to the accountability requirements.

5.2 PLOS ONE (2)

None of the co-authors of the paper titled “**The ^1H -NMR-based metabolite profile of acute alcohol consumption: a metabolomics intervention study**” had any objection to the use of the paper, as presented in Chapter 4 (section 4.5.2), as part of the thesis submitted for the degree *Philosophiae Doctor*.

The contributions of the co-authors for this publication were as follows: Prof Carolus Reinecke and Prof Johan Westerhuis defined the aim of the study, developed the methodological design and supervised the planning and execution of the investigation. Prof Carolus Reinecke also obtained research funds for the project, ethical approval from the Health Sciences Ethical Committee of North-West University, and written informed consent from the 24 participants of the cross-over study. Ms Cindy Irwin conducted all experimental analyses and data collection. Dr Shayne Mason gave guidance and assessment on ^1H -NMR data generation and spectral analyses. Ms Cindy Irwin and Dr Shayne Mason were responsible for the data curation, peak annotation and production of the metadata. Ms Mari van Reenen and Prof Johan Westerhuis conceptualized the complementary techniques required to analyse the data and Ms Mari van Reenen applied selected computer programmes, implemented computer codes and supporting algorithms, and performed the statistical and computational analyses of the original and curated data. Prof Lodewyk Mienie and Prof Ron Wevers participated throughout the investigation as expert on metabolism and in modelling of the proposed metabolic outcome of the study. All authors gave input according to personal expertise in the development of the manuscript, gave final approval of the version as submitted and agreed to the accountability requirements.

5.3 Scientific Reports

None of the co-authors of the paper titled **“GC–MS-based urinary organic acid profiling reveals multiple dysregulated metabolic pathways following experimental acute alcohol consumption”** had any objection to the use of the paper, as presented in Chapter 5 (section 5.7), as part of the thesis submitted for the degree *Philosophiae Doctor*.

The contributions of the co-authors for this publication were as follows: Prof Carolus Reinecke and Prof Johan Westerhuis were responsible for conceptualization of the study, development of the methodological design, and administration of the project. Prof Carolus Reinecke also obtained research funds and ethical approval for the project, and supervised the planning and execution of the investigation. Ms Cindy Irwin conducted all experimental analyses and data collection. Ms Cindy Irwin, Dr Shayne Mason and Prof Lodewyk Mienie were involved in the GC–MS data generation, data curation and analysis. Ms Mari van Reenen and Prof Johan Westerhuis conceptualized the complementary techniques required to analyse the data and Ms Mari van Reenen applied selected computer programmes, implemented computer codes and supporting algorithms, and performed the statistical and computational analyses of the original and curated data. Prof Lodewyk Mienie and Prof Ron Wevers participated throughout the investigation as experts on metabolism, and in modelling of the proposed metabolic outcome of the study. All authors gave input according to personal expertise in the development and writing of the manuscript, gave final approval of the version as submitted, and agreed to the accountability requirements.

6 Copyright licences for figures

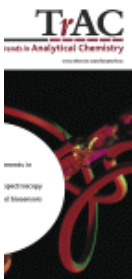
6.1 **Figure 2–1:** Schematic representation of the conversion between NAD^+ and NADH during redox reactions (adapted from Rye *et al.*, 2016: “Download for free at <https://openstax.org/details/books/biology>”).

Biology by OpenStax is licensed and distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

©2017 Rice University. Textbook content produced by OpenStax is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Under this license, any user of this textbook or the textbook contents herein must provide proper attribution as follows:

- If you redistribute this textbook in a digital format (including but not limited to PDF and HTML), then you must retain on every page the following attribution: “Download for free at <https://openstax.org/details/books/biology>.”
- If you redistribute this textbook in a print format, then you must include on every physical page the following attribution: “Download for free at <https://openstax.org/details/books/biology>.”
- If you redistribute part of this textbook, then you must retain in every digital format page view (including but not limited to PDF and HTML) and on every physical printed page the following attribution: “Download for free at <https://openstax.org/details/books/biology>.”
- If you use this textbook as a bibliographic reference, please include <https://openstax.org/details/books/biology> in your citation.

6.2 Figure 3–1: Metabolomics workflow (adapted with permission from Hendriks *et al.*, 2011)



Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Ms. Cindy Irwin
Order Date:	Oct 25, 2017
Order Number:	4215891331217
Publication:	TrAC Trends in Analytical Chemistry
Title:	Data-processing strategies for metabolomics studies
Type of Use:	reuse in a thesis/dissertation
Order Total:	0.00 USD

View or print complete details of your order and the publisher's terms and conditions

<https://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=1a38a0d8-bb88-4b97-9ba2-f9407d1aa456>

Tel: [+1-855-239-3415](tel:+1-855-239-3415) / [+1-978-646-2777](tel:+1-978-646-2777)
customercare@copyright.com
<https://myaccount.copyright.com>



6.3 Figure 3–2: OPLS score plot and spectra of chronic liver failure (CLF) and acute-on-chronic liver failure (ACLF) patients (Amathieu *et al.*, 2014).

Copyright: © 2014 Amathieu *et al.* This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

6.4 Figure 5–1: Number of NMR metabolomics and MS metabolomics publications (reproduced with permission from Emwas, 2015).



6.5 Figure 5–3: Representation of all elements of the experimental design (Irwin *et al.*, 2016).

Copyright: © 2016 Irwin *et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.