



TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Experimentelle Genetik

**Metabolomic study of alcohol consumption – Investigation
of novel candidate biomarkers and their potential clinical
use**

Marta Jaremek

Vollständiger Abdruck der von der Fakultät für Wissenschaftszentrum
Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen
Universität München zur Erlangung des akademischen Grades eines Doktors der
Naturwissenschaften genehmigten Dissertation.

Vorsitzende(r): Prof. Dr. Heiko Witt
Prüfer der Dissertation: 1. Prof. Dr. Jerzy Adamski
2. Prof. Dr. Johann J. Hauner

Die Dissertation wurde am 22.05.2018 bei der Technischen Universität München
eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für
Ernährung, Landnutzung und Umwelt am 01.08.2018 angenommen.

Zusammenfassung

Alkoholkonsum kann die Gesundheit des Einzelnen schädigen und erhebliche sozio-ökonomische Schäden verursachen. Klinische Tests zur genauen Bestimmung des Alkoholkonsums als auch Biomarker für die frühzeitige Erkennung alkoholbedingter Erkrankungen sind jedoch limitiert und die zugrundeliegenden biochemischen Mechanismen noch weitgehend unbekannt. Seit den 1980er Jahren haben nur wenige Tierstudien den Zusammenhang zwischen Alkohol und der metabolischen Veränderungen untersucht. Das Ziel dieser Arbeit war es, das Wissen im Bereich der Alkohol-Biomarker-Discovery zu erweitern um neue Erkenntnisse über Alkohol induzierte metabolische Veränderungen beim Menschen zu gewinnen und daraus neue potentielle klinische Alkoholbiomarkerkandidaten abzuleiten. Da Alkoholkonsum eine Vielzahl von Organen, biologischen Prozessen und Signalwegen beeinflusst, ist ein holistischer Biomarker-Discovery-Ansatz, der näher am Phänotyp ist und einen globalen und direkten Einblick in die Entwicklung und das Fortschreiten von Alkohol induzierter Toxizität gibt, dringend erforderlich. Im Rahmen einer ersten Targeted Metabolomics Alkohol Studie, wurde die neue Hochdurchsatz-Tandem-Massenspektrometrie-Plattform in einer großen europäischen Bevölkerungsgruppe eingesetzt um metabolische Effekte von Alkoholkonsum auf den menschlichen Organismus zu untersuchen. Die Studie identifizierte Alkohol assoziierte Metaboliten und ihre potenziell zugrunde liegenden Mechanismen. Die Studie überbrückt die Wissenslücke von Alkohol und seiner Auswirkungen auf das Metabolom und stellt eine zusätzliche Verbindung her im Verständnis der schrittweisen Entwicklung vom chronischen moderaten-bis-hohen Alkoholkonsum bis hin zu seinen potentiellen klinischen Endpunkten.

Abstract

Alcohol consumption can harm the health of individuals and cause significant socio-economic damage. However, clinical tests for the accurate assessment of alcohol consumption and biomarkers for the early detection of alcohol-related diseases are limited and the underlying biochemical mechanisms are still not fully understood. Since the 1980s, only few animal studies have investigated the association between alcohol and metabolic changes. The aim of this work was therefore to expand the knowledge in the field of alcohol biomarker discovery to gain new insights into alcohol-induced metabolic changes in humans and possibly to derive new clinical alcohol biomarker candidates. Since alcohol consumption affects a variety of organs, biological processes and signaling pathways, a holistic biomarker discovery approach that is closer to the phenotype and that provides a global and direct insight into the evolution and progression of alcohol-induced toxicity is urgently needed. As part of a first Targeted Metabolomics alcohol study, the new high-throughput tandem mass spectrometry platform was deployed in a large European population to study the metabolic effects of alcohol consumption on the human organism. The study identified alcohol-associated metabolites and their potentially underlying mechanisms. The study bridges the knowledge gap of alcohol and its effects on the metabolome and provides an additional link in understanding the gradual evolution from chronic moderate-to-high alcohol consumption to its potential clinical endpoints.

Contents

Zusammenfassung	2
Abstract.....	3
1. Introduction	8
1.1 Origins of alcohol consumption	8
1.1.1 Evolutionary aspects	8
1.1.2 Historical development of alcohol consumption in society	8
1.2 Alcohol consumption – a health issue today	9
1.3 Need for new biomarkers in alcohol diagnostics and treatment.....	11
1.3.1 Socio-economic consequences of alcohol-related disorders	11
1.3.2 Current diagnostic shortcomings	11
1.4 Metabolomics - the new approach in biomarker discovery	14
1.4.1 The next generation “Omics”	14
1.4.2 Mass-spectrometry as an enabling high-throughput technology.....	15
1.5 Current state of research - Studies on alcohol-related metabolic perturbations	18
2. Objectives	24
2.1 Identification of candidate alcohol biomarkers	24
2.2 Biochemical and functional interpretation.....	24
2.3 Clinical utility	24
3. Materials and Methods	25
3.1 Baseline research and literature reviews	25
3.2 Metabolomics study design	25
3.3 Study populations	25
3.3.1 KORA F4 discovery data set	25
3.3.2 KORA F3 replication data set.....	27

3.3.3	TwinsUK replication data set	27
3.3.4	Data collection	27
3.3.5	Assessment of alcohol consumption	27
3.3.6	Ethics Statement.....	28
3.4	Blood sampling	28
3.5	Metabolite measurements	29
3.6	Data pre-processing	30
3.6.1	Outlier detection.....	30
3.6.2	Missing value imputation	30
3.6.3	Data normalization.....	31
3.6.4	Data distribution check	31
3.7	Statistical data analysis	31
3.7.1	Logistics regression analysis.....	31
3.7.2	Random forest selection	32
3.7.3	Stepwise selection	32
3.7.4	Akaike's Information Criterion (AIC).....	33
3.7.5	Correlation analysis	33
3.7.6	Area under the receiver-operating characteristic (AUROC)	33
3.7.7	Clustering and Heat maps	34
3.7.8	Meta-analysis.....	34
3.7.9	Smoother plots.....	34
3.8	Metabolite biochemical and functional Interpretation	35
3.8.1	Metabolite biochemical interpretation	35
3.8.2	Metabolic Pathway and Functional Enrichment analysis	35
3.8.3	Literature research on alcohol-related metabolites and associated disease states and phenotypes	36

4.	Results	36
4.1	Baseline characteristics of the discovery and replication data sets (KORAF4, KORAF3 and TwinsUK).....	36
4.1.1	Drinkers	36
4.1.2	Abstainers.....	37
4.2	Global metabolic differences between Abstainers, Low-drinkers and Moderate-to-heavy drinkers.....	40
4.3	Cluster analysis of significant global metabolic changes in Low-Drinkers and Moderate-to-Heavy drinkers	48
4.4	Identification of alcohol-related key metabolites.....	49
4.4.1	Metabolite selection	49
4.4.2	Model evaluation.....	50
4.5	Replication analysis in two independent cohorts.....	51
4.6	Characterization of the relationship between alcohol consumption and key metabolite concentrations	55
4.7	Correlation of key metabolites and laboratory parameters (HDL and triglycerides)	58
4.8	Alcohol consumption effects on metabolic pathways	60
4.9	Alcohol-related metabolites and associated disease states & phenotypes	62
5.	Discussion.....	64
5.1	Metabolomics study	64
5.1.1	Study design	64
5.1.2	Data quality.....	64
5.1.3	Study populations (Alcohol consumption groups)	64
5.1.4	Methodology to identify significant metabolites	65
5.1.5	Metabolic Pathway analysis	65
5.2	Alcohol-related metabolites	66

5.2.1	Metabolomic candidate biomarker panel and its potential alcohol-related underlying biochemical mechanisms.....	66
5.2.2	Alcohol-related metabolites and their association with medical conditions.....	71
5.3	Clinical utility of metabolic candidate biomarkers.....	73
5.3.1	Test scenarios in alcohol-related investigations.....	73
5.3.2	Potential application areas	74
5.3.3	Mass spectrometry in the clinical lab.....	75
5.3.4	What's next? - alcohol metabolomics study in the context of the biomarker development workflow	83
6.	Conclusion and Perspectives	87
6.1	Metabolomics study of alcohol intake.....	87
6.2	Value of mass spectrometry for laboratory diagnostics	87
6.3	Perspectives in alcohol metabolomics-based research	87
6.4	Contribution of the underlying work to alcohol research	88
	References	90
	List of Figures	100
	List of Tables	102
	Publications	103
	Danksagung	104

1. Introduction

1.1 Origins of alcohol consumption

1.1.1 Evolutionary aspects

How did alcohol come into being? - according to the American Biologist Robert Dudley (Dudley 2014) alcohol consumption *did not suddenly fall from the sky*, but has evolved over millions of years. About 2.5 million years ago until about 10 000 years ago, our phylogenetic ancestors have consumed alcohol eating ripe fruits. The localization of those fruits has been associated with high-energy foods and provided an evolutionary benefit to our ancestors. This type of alcohol consumption has shaped the behavior of primates, who with time have adjusted to the intake of low ethanol concentrations. The theory of evolutionary origin of alcohol consumption is supported by the heritability as well as the biogeographic distribution of alcohol dehydrogenase, the most important enzyme for alcohol degradation in humans (Dudley 2014).

1.1.2 Historical development of alcohol consumption in society

Drinking of alcohol (e.g. wine and beer) is known since in approx. 5.000 BC the inhabitants of Mesopotamia have invented the manufacturing of alcoholic beverages. Later, during the Middle Ages, discovery and development of alcohol distillation made it possible to produce alcoholic beverages with an ethanol concentration of up to 80 volume percent. During the course of history, alcohol had different functions and consequences across different cultures. Whereas in the Middle Ages the function of alcohol was mainly food (i.e. beer), stimulant, facilitation of social contacts, sacral substance, narcotic and medication, the focus of alcohol in the Industrialization period shifted with time from functions to consequences such as an addictive substance, cause of social problems and physical damage (Feuerlein 1994)

1.2 Alcohol consumption – a health issue today

Today, alcohol still plays an important role in everyday life. Notably, it is an extraordinary substance, as unlike other substances (i.e. tobacco) it is psychoactive, addictive and at the same time it can be consumed without legal sanctions except for a few restrictive exceptions ((AWMF) 2016). Alcohol follows the pharmacological leitmotif "*Dosis sola venenum facit*" (i.e. only the dose makes the poison). In small quantities, alcohol can have a number of positive effects. However, this is not intended to obscure the fact that acute intoxication can occur in the case of increased alcohol consumption, and multiple organ damage can occur in the case of chronic intake of larger amounts of alcohol (Ashley, Rehm et al. 2000), (World Health Organization 2014). There is hardly an organ that cannot be harmed by alcohol (**Figure 1**). Alcohol consumption has been identified as a component cause for more than 200 diseases, injuries and other health conditions with ICD-10 codes (World Health Organization 2014), (Shield, Parry et al. 2013). The individual organs, however, are often affected with varying frequency and individual severity with comparable alcohol consumption. The underlying mechanisms are still largely unexplained. Therefore, research on the effect of alcohol on human body and especially the early detection of alcohol-induced somatic and mental health conditions in asymptomatic individuals is of great importance

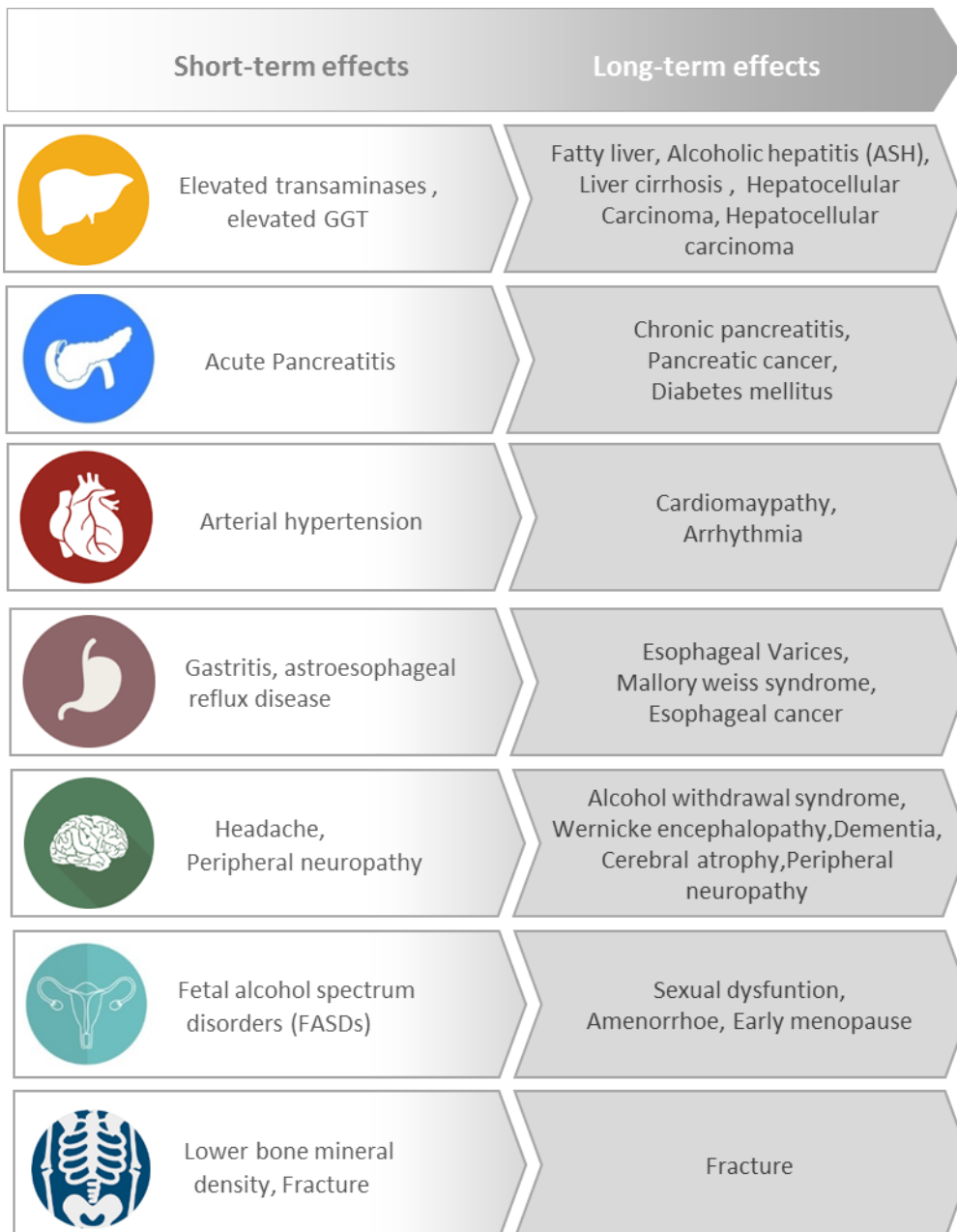


Figure 1. Main short-term and long-term effects of alcohol consumption. Modified according to <https://www.unitypoint.org> and <https://www.help4addiction.co.uk/effects-alcohol-body/>.

1.3 Need for new biomarkers in alcohol diagnostics and treatment

1.3.1 Socio-economic consequences of alcohol-related disorders

Alcohol-induced disorders are common, frequently underdiagnosed (as patients may not disclose alcohol consumption or underreport alcohol consumption) and expensive at total costs (Moore, Bone et al. 1989). The estimated costs of alcohol-related disorders are at 180 billion USD in the US, 125 billion Euros in Europe and 24,4 billion Euros in Germany (dhs 2009). Costs are caused by loss of production due to illness, accident and death, treatment costs and damage to property. The global cost for the sole treatment of alcohol-related diseases cannot be easily fully estimated. However, as an example, according to 2009 figures in the UK, the National Health Service (NHS) spends an estimated 3 billion British Pounds per year treating alcohol-related illness (Gika and Wilson 2014).

1.3.2 Current diagnostic shortcomings

Traditional biomarkers are not alcohol specific

In the sense of an early diagnosis and therapy of alcohol-related disorders, questionnaires such as CAGE (Ewing 1984) or the Alcohol Use Disorder Identification Test (AUDIT) (Saunders, Aasland et al. 1993) as well as biomarkers can be used. Alcohol biomarkers are physiological indicators of alcohol exposure (SAMHSA 2012). Indirect biomarkers suggest heavy alcohol consumption by detecting the endogenous toxic effects that alcohol may have had on organ systems or body chemistry (Litten, Bradley et al. 2010). However, the current indirect traditional biomarkers used in clinical routine, such carbohydrate deficient transferrin (CDT), gamma-glutamyl-transferase (GGT) or mean corpuscular volume (MCV), have limitations as they are not alcohol-specific and are influenced by other

substances and by non-alcohol-related disorders (Conigrave, Degenhardt et al. 2002, Allen, Marques et al. 2008), (Hannuksela, Liisanantti et al. 2007), (Helander 2003), (Laposata 1999), (Niemela 2007) (Torrente, Freeman et al. 2012).

Most biomarkers exhibit a narrow window of alcohol detectability

In dealing with alcohol-related human toxicity and disease, there is a need for a range of different types of biomarkers. For example, because of the relatively short PK half-life of ethanol, markers are needed that indicate long-term exposure to help monitor therapy and rehabilitation (NIH 2006). Currently only the MCV and FAEE (fatty acyl ethyl esters) can remain reliably measurable for a long time in the body after drinking is stopped (**Figure 2**).

Direct ethanol metabolites provide no information on progression of toxicity

In addition to indirect biomarkers, direct ethanol metabolites are increasingly being observed in the last decade. They are termed “direct” because they are analytes of alcohol breakdown. When alcohol is consumed, the biggest portion is metabolized by the oxidative processes in the liver, but a small amount is also broken down non-oxidatively, creating analytes that can be measured in blood or urine. At present, the following direct ethanol metabolites are used ethyl glucuronide (EtG), ethylsulfate (EtS), fatty acid ethyl ester (FAEE), phosphatidylethanol (PEth) (Foti and Fisher 2005), (Helander and Beck 2004), (Kaphalia, Cai et al. 2004), (Varga, Hansson et al. 1998) These are not biomarkers in themselves, but metabolites of the toxin, and while useful in demonstrating/quantifying alcohol exposure, do not necessarily, by themselves, provide information on the mechanisms of toxicity.

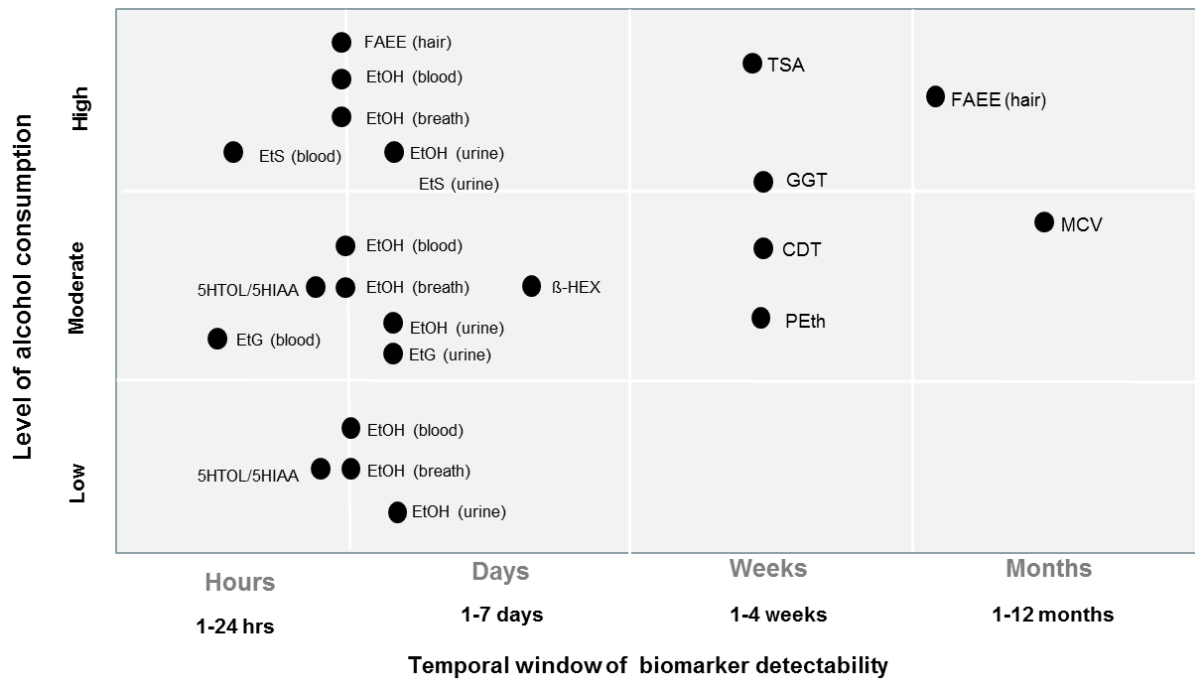


Figure 2. Current traditional alcohol biomarkers are limited. One of the diagnostic shortcomings today is that most current biomarkers exhibit a narrow window of alcohol detectability. Y-axis: estimated point in time biomarker remains reliably measurable in the body after drinking is stopped, x-axis : level of lowest alcohol consumption required to induce biomarker changes compared to reference range; exact definitions of low, moderate and high alcohol consumption varies across publications. Schematic modified according to <http://pubs.niaaa.nih.gov/publications/arh341/56-63.htm>

Exposure to alcohol affects a variety of organs, biochemical and signaling pathways, and biological process. Thus, conventional deductive biology, which provides only a fragmented view of a complex picture, is arguably not up to the challenge of understanding the metabolic changes of long-term alcohol exposure. Biomarkers that give insights into the development and progression of toxicity are required, which might be best served using a holistic “systems biology” approach (Gika and Wilson 2014). The comprehension of such biological complexity is best achieved via exploratory/and or holistic approaches, known in the modern biology as “Omics” sciences (Micheel 2012).

1.4 Metabolomics - the new approach in biomarker discovery

1.4.1 The next generation “Omics”

Since the 1990s, biomarker research was mainly focused on the areas of Genomics, Transcriptomics and Proteomics. However, during the last couple of years the full complexity of molecular biology has been realized and the complex interactions between genetic make-up and especially environmental factors have now been recognized. It is now accepted that understanding of these interactions is impossible at only genetic, transcriptomic and proteomic level. Metabolomics is a rather recent concept in biomarker research that aims to quantify all metabolites in a biological system (Fiehn 2002). Metabolomics is based on a holistic approach for the characterization of metabolites present in biological samples (i.e. endogenous metabolites). Focus of analysis is endogenous metabolites with a molecular mass of 80 – 1,200 Da (Hocher and Adamski 2017). Unless the other “Omics” approaches as Genomics, Transcriptomics or Proteomics, Metabolomics applications measure metabolites to provide a direct assessment of human health and its influences i.e. environmental factors like nutrition, lifestyle, smoking, exercise and medication. Due to its close proximity to the phenotype, metabolic biomarkers are identifiable with real biological endpoints of a biological system. Numerous publications have already shown successful applications of metabolomics in biomarker discovery (Xu, Holzapfel et al. 2013),(Mittelstrass, Ried et al. 2011, Wang-Sattler, Yu et al. 2012, Floegel, Wientzek et al. 2014, Vouk, Ribic-Pucelj et al. 2016). Metabolomics studies are divided into two different types : targeted and non-targeted approaches : non-targeted metabolomics is used for a global metabolite analysis, that is, comprehensive analysis of all the measurable analytes in a sample (including analyte identification of unknown signals) (Griffiths, Koal et al. 2010), whereas in a targeted

metabolomics strategy, predefined metabolite-specific signals are often used to determine precisely and accurately relative abundancies and concentrations of limited number of metabolites. Targeted metabolomics approach is typically applied in hypothesis-driven studies with focus on prediction of metabolic differences of known metabolites between cohorts of populations and for understanding of the underlying regulatory mechanism of metabolic pathways (Fiehn 2007) .

1.4.2 Mass-spectrometry as an enabling high-throughput technology

Mass spectrometry is the key technology for metabolomics. Mass spectrometry is an analytical technique to determine the mass-to-charge ratio (m/z) of particles (Nair and Clarke 2017) (**Figure 3**). The underlying platforms consist of different instrument configurations (**Figure 4**). It is a rapidly developing technology for both qualitative and quantitative analyses of proteins and peptides (Popp, Malmstrom et al. 2015) oligonucleotides (Wolk, Kaleta et al. 2012), drug metabolites (Desrosiers, Scheidweiler et al. 2015), steroids (Haller, Prehn et al. 2010), volatile organic compounds (VOC) (Jaremek and Hiltawsky 2012) and endogenous metabolites as acyl carnitines or amino acids (Millington, Kodo et al. 1990). As of today, the pressure is on for high-throughput approaches to accelerate the generation, identification and optimization of biomolecules. As traditional methods of analysis become antiquated, new analytical strategies and techniques are necessary to meet sample throughput requirements and man power constraints. Recent advancements in tandem mass spectrometry have made tremendous impact on the identification of components in Newborn screening applications (Chace, DiPerna et al. 1999), Therapeutic Drug Monitoring (TDM) (McShane, Bunch et al. 2016) and Drugs of Abuse (DoA) (Eichhorst, Etter et al. 2009) directly from human fluid. In this respect, many clinical laboratories in both academia and industry are taking advantages of

mass spectrometry developments.

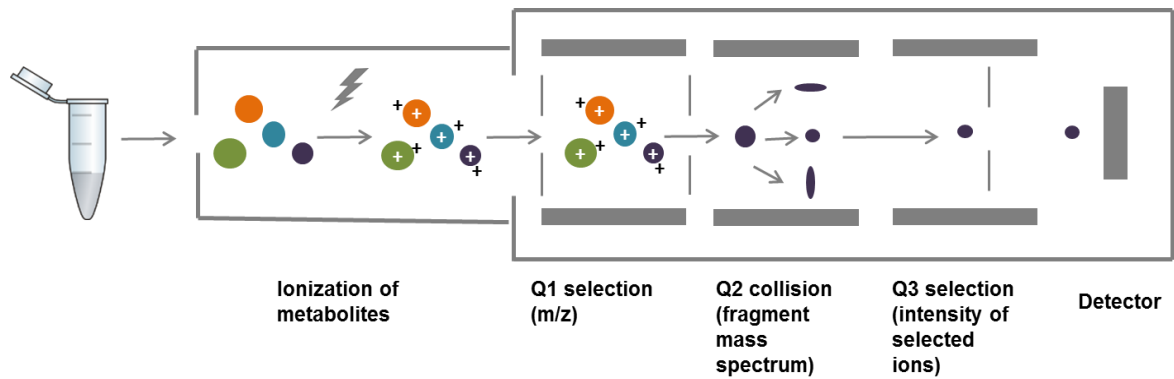


Figure 3. Mass spectrometry technology principle. Key element of mass spectrometry is the ionization of analytes and their subsequent separation according to their mass-to-charge ratio (m/z). The mass spectrometer can be constructed in different ways, here the graphic depicts a tandem mass spectrometer; metabolites (displayed as colored bubbles) are extracted from the sample (e.g. using a chromatographic step) and then applied to ionization chamber. Ionized molecules are pre-selected by m/z ratio in the first quadrupole Q1. In the second quadrupole (Q2) molecules released from Q1 are fragmented and applied to third quadrupole (Q3). In Q3 fragments of molecules of interest are selected and quantified in the detector. Modified according to (Hoche and Adamski 2017)

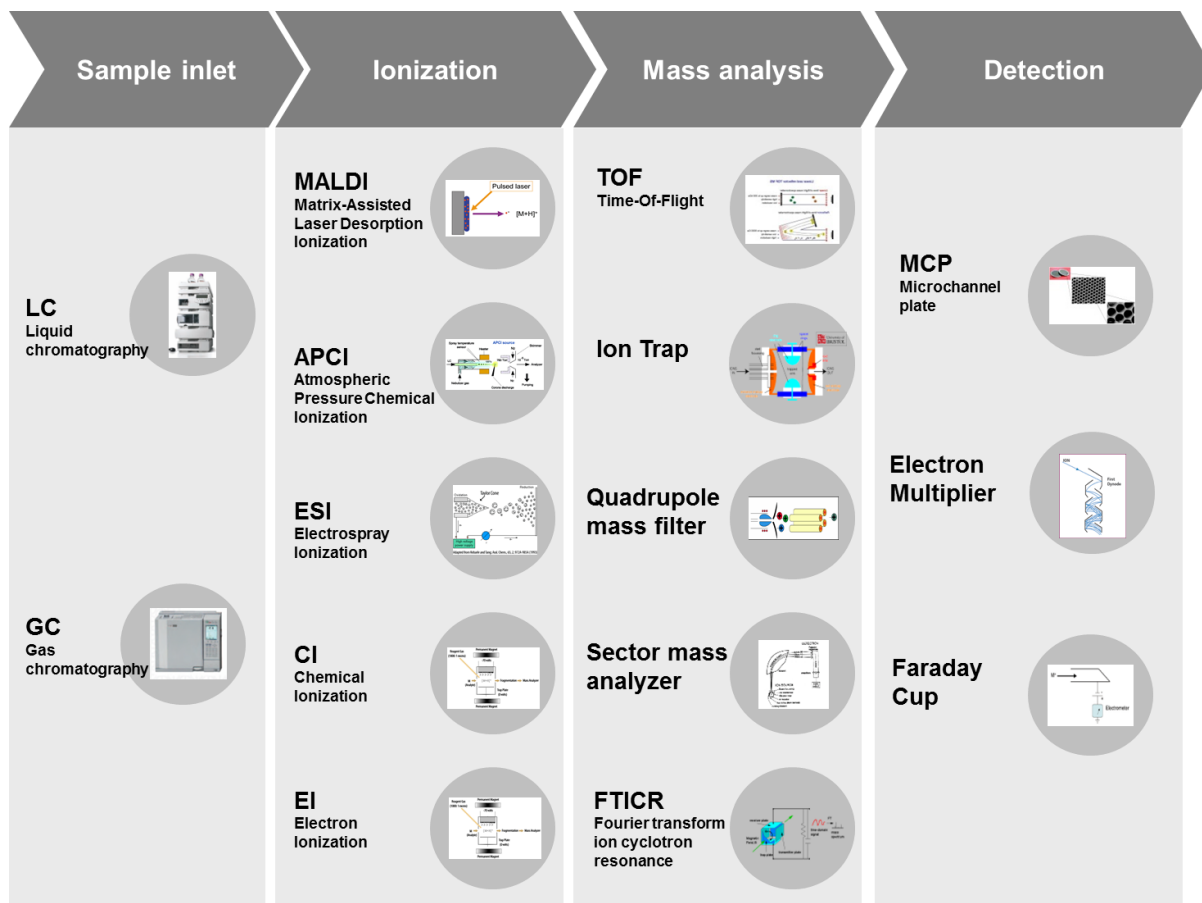


Figure 4. Common technologies and configurations of mass spectrometry instrumentation. The main components of a mass spectrometer are an Inlet system e.g. (LC, GC, direct probe etc.), Ion source (e.g. MALDI, APCI, ESI, CI, EI, etc.), Mass analyzer (TOF, Ion Trap, Quadrupole, Sector mass analyzer, etc.) and Detector (e.g. Micro Channel Plates, Electron Multiplier, etc.). Depending on the sample's chemical and physical properties, different ionization techniques can be used. For example, for samples that are not thermolabile and relatively volatile, ionization such as Electron Impact and/or Chemical Ionization is typically used. For samples that are thermolabile such as peptides, proteins and other samples of biological interest, soft ionization techniques are to be considered. Schematic compiled according to mass spectrometry segmentation tutorial materials from (Iowa State University 2017).

1.5 Current state of research - Studies on alcohol-related metabolic perturbations

Studying the metabolic processes that lead to gradual development of disease provides an excellent way of discovering biomarkers of toxicity and for determining the extent of damage to the organism. Despite the importance of ethanol in causing human disease and the limitation of traditional alcohol biomarkers, relatively few original research studies using metabolomics platforms have been performed to date (**Table 1**). Those studies that have been undertaken have largely been in rodents except one study using human embryonic stem (hES) cells (Palmer, Poenitzsch et al. 2012) and one study using 61 human males (Vazquez-Fresno, Llorach et al. 2012). The technique that has been most applied in metabolic studies on alcohol research in the case of ethanol is LC-MS, particularly in combination of UPLC with TOF followed by H¹-NMR with a single study that employed GC-MS. Most of the published work on metabolomics for the study of alcohol related metabolic dysfunction relate to the analysis of urine or liver samples of rodents, followed by reports describing metabolic studies on brain, serum, plasma and pancreatic tissue. In most of studies in rodents, the alcohol was administered via oral gavage or in the animals drinking water. The literature reports studies on both acute and chronic alcohol consumption in rodents, with animal treated with ethanol after single dose or exposed for periods ranging from 2 weeks to 6 months. Almost all these reports are supported by additional methodologies with clinical parameters measurements, histopathology data on liver tissues or proteomics analyses. In all studies a relatively small number of samples was studied, in some studies the number of samples was not defined. Also, a relatively small number <30 metabolites was measured. (Weiner, Coker et al. 1984, Bradford, O'Connell et al. 2008, Nicholas, Kim et al. 2008, Wang, Lv et al. 2008,

Masuo, Imai et al. 2009, Fernando, Kondraganti et al. 2010, Manna, Patterson et al. 2010, Fernando, Bhopale et al. 2011, Li, Liu et al. 2011, Loftus, Barnes et al. 2011, Manna, Patterson et al. 2011, Gika, Ji et al. 2012, Palmer, Poenitzsch et al. 2012, Shi, Yao et al. 2012, Vazquez-Fresno, Llorach et al. 2012, Liu, Raghu et al. 2013, Yoseph, Breed et al. 2013, Fan, Cao et al. 2014)

With the advent of the latest targeted high-throughput metabolomics technology by Biocrates (BIOCRATES Life Sciences AG, Innsbruck, Austria) and the availability to measure >300 known metabolites in one run, the research question arose to investigate alcohol-induced metabolic profiles in a big European human population to derive new insights for alcohol research and potentially identify novel biomarkers of alcohol consumption.

Reference	Phenotype/ Disease state under study	Outcome	Metabolomics- based method	Sample <i>(for metabolomics analysis)</i>	Technology	Instrumentation configuration	Candidate Biomarkers identified/ Significant metabolite changes <i>(as described in the publication)</i>
Weiner et al. 1984	Relationship between alcohol consumption and metabolic profiles	Changes in metabolism produced by alcohol	Metabolic profiling	urine	MS	GC/MS	Threonic acid; Glucuronic; an undetermined acid (No. 55); pyroglutamic acid
Nicholas et al. 2007	Relationship between acute and chronic alcohol fed and metabolic levels	Both acute and binge ethanol caused (i) decreased glucose, lactate, and alanine in liver and serum; (ii) increased acetate in liver and serum; and (iii) increased acetoacetate in serum. Binge-ethanol increased liver β -hydroxybutyrate and decreased betaine.	Metabolomics	liver tissue serum	NMR	1H-NMR	Acute and Binge alcohol consumption : Glucose, Lactate, Alanine ,Acetate, Acetoacetate Binge alcohol consumption : β -hydroxybutyrate Betaine
Bradford et al 2008	Steatohepatitis Liver disease	Changes in N-acetylglutamine and n-acetylglycine may be useful markers of alcohol consumption	Metabolomic profiling	urine liver	NMR MS	ESI/FTICR-MS NMR	N-acetylglutamine n-acetylglycine
Wang et al 2008	Relationship between alcohol-induced hepatotoxicity and metabolite levels	Sphingomyelin signaling pathway is involved in alcohol hepatotoxicity	Metabolic profiling	urine	MS	UPLC/ESI- QTOF-MS	ceramide (d18:1/25:0)
Masuo et al 2009	Relationship between Sake consumption and metabolic profiles	Significant increases in valine, arginine/ ornithine, alanine, glutamine, and choline with decreases in isoleucine, N-acetyl aspartate, taurine, glutamate, and gamma aminobutyric acid in brain	Metabolomics	brain, liver tissue	NMR	1H-NMR	Valin arginine/ ornithine, alanine, glutamine, and choline with decreases in isoleucine, N-acetyl aspartate, taurine, glutamate, and gamma aminobutyric acid in brain
Fernando et al 2010	Hepatic steanosis Alcoholic liver disease Fatty Liver	Potential method to detect early-stage-alcohol-induced fatty liver disease by analyzing plasma lipid profile	Lipidomics	plasma liver tissue	NMR	1H-NMR 31P-NMR	TG Phospholipids

Table 1. Current studies on alcohol-induced metabolic perturbations. Shows the result of literature baseline analysis using PubMed.

Reference	Phenotype/ Disease state under study	Outcome	Metabolomics- based method	Sample <i>(for metabolomics analysis)</i>	Technology	Instrumentation configuration	Candidate Biomarkers identified/ Significant metabolite changes <i>(as described in the publication)</i>
Manna et al 2010	Alcohol-induced Liver Disease (ALD)	Demonstration of metabolomics approach to identify early, noninvasive biomarkers of ALD pathogenesis	metabolomics	urine	MS	UPLC- ESI/TOF- MS	Ethyl-sulfate Ethyl-β-D-glucuronide 4-hydroxyphenylacetic acid 4-hydroxyphenylacetic acid sulfate 2-hydroxyphenylacetic acid adipic acid pimelic acid indole-3-lactic acid
Fernando et al 2011	Alcohol Liver Disease (ALD)	n.a.	Lipidomics	plasma liver tissue	NMR	1H-NMR 31P-NMR	Phosphatidylcholine Lysophosphatidylcholine Phosphatidylserine Phosphatidylethanolamine Lyso-Phosphatidylethanolamine
Li et al 2011	Liver injury Hepatocellular carcinoma (HCC)	LPC profiles in serum may be biomarkers for liver injury and hepatocellular carcinoma	Metabolomics	serum	MS	UPLC/Q-TOF MS	Leucine; Phenylpyruvic acid; Phenylalanine Tryptophan; LPE (16:0); LPE (18:0) Lysophosphatidylcholines : LPC (16:0) LPC (18:0) LPC (20:1) LPC (22:6) Phosphatidylcholines : PC (16:0/18:3) PC (12:1/24:3) PC (16:0/20:4) PC (16:0/22:6) PC (16:0/22:6) PC (18:0/20:4) Sphingomyelins : SM (d18:0/16:1)

Table 1 (continued). Current studies on alcohol-induced metabolic perturbations. Shows the result of literature baseline analysis using PubMed

Reference	Phenotype/ Disease state under study	Outcome	Metabolomics- based method	Sample <i>(for metabolomics analysis)</i>	Technology	Instrumentation configuration	Candidate Biomarkers identified/ Significant metabolite changes <i>(as described in the publication)</i>
Loftus et al 2011	Alcoholism	Increased metabolite levels in livers in rat and mouse by alcohol treatment.	Metabolite profiling	liver tissue	MS	LCMS-IT-TOF	Fatty acyls metabolites: octadecatrienoic acid eicosapentaenoic acid Fatty acid ethyl esters (FAEE)s : ethyl arachidonate ethyl docosahexaenoic acid ethyl linoleate ethyl oleate Phosphatidylethanol (PEth) homologues : PEth 18:0/18:2 PEth 16:0/18:2
Manna et al 2011	Alcohol-induced Liver Disease (ALD)	Indole-3-lactic Acid and Phenyllactic Acid are potential candidates for conserved and pathology-specific for early stages of ALD	metabolomics	urine	MS	UPLC-ESI/TOF-MS	Indole-3-lactic Acid Phenyllactic Acid
Gika et al 2012	Relationship between chronic alcohol consumption and metabolic profiles	A large number of metabolites were seen to differ between control and alcohol-treated animals, for both biofluids.	Metabolic profiling	plasma, urine	MS	UHPLC-TOF MS	Octadecatrienoic acid Eicosapentaenoic acid Ethyl arachidonate Ethyl docosahexaenoic acid Ethyl linoleate Ethyl oleate Phosphatidylethanol (PEth) homologues (including PEth 18:0/18:2 and PEth 16:0/18:2)
Palmer et al 2012	Relationship between ethanol exposure and metabolite levels of human embryonic stem cells	Ethanol exposure induces changes to metabolite profile of human embryoid bodies	Metabolomics	human embryonic stem (hES) cells	MS	ESI-QTOF-MS	L-thyroxine 5'-methylthioadenosine tryptophan metabolites L-kynurenine indoleacetaldehyde
Shi et al 2012	Relationship between Ethanol and Hepatic Steatosis	Metabolomic analysis of urinary metabolites revealed time- and dose-dependent changes in the chemical composition of urine.	Metabolomics	urine	MS	ESI-QTOF-MS	4-Hydroxyphenylacetic acid sulfate; Ethyl glucuronide (EtG); Ethyl sulfate (EtS) N-Acetylneuraminic acid

Table 1 (continued). Current studies on alcohol-induced metabolic perturbations. Shows the result of literature baseline analysis using PubMed

Reference	Phenotype/ Disease state under study	Outcome	Metabolomics- based method	Sample <i>(for metabolomics analysis)</i>	Technology	Instrumentation configuration	Candidate Biomarkers identified/ Significant metabolite changes <i>(as described in the publication)</i>
Vazquez-Fresno et al 2012	Relationship between moderate wine consumption and metabolite levels	Ability of metabolomics to obtain metabolomic picture including food metabolome and endogenous biomarkers of moderate wine intake	Metabolomics	urine	NMR	H-NMR	Branched-chain amino acid (BCAA) 3-methyl-oxovalerate
Liu et al 2013	Alcohol-induced Fatty Liver Disease (AFLD)	Increased metabolite levels after alcohol administrations. Levels of metabolites recovered in ginger treatment group	Metabolic profiling	serum	MS	HPLC/QTOF-MS	D-glucurono-6,3-lactone; glycerol-3-phosphate; pyruvic acid; lithocholic acid; 2-pyrocatechuic acid; prostaglandin E1
Yoseph et al 2013	alcohol-induced mortality and morbidity	Septic alcohol-fed mice had a significantly higher mortality than septic water-fed mice	¹ H NMR metabolomics	pancreatic tissue	NMR	¹ H NMR	Acetate; Adenosine; Xanthine; Acetoacetate; 3-Hydroxybutyrate; Betaine Cytidine; Uracil; Fumarate; Creatine;Phosphate; Creatine;Choline
Fan et al 2014	Relationship between chronic alcohol consumption and changes in heart and serum metabolite profiles	Low to moderate alcohol consumption increases HDL cholesterol	Metabolomics	serum heart tissue	MS	HPLC-TOF-MS	HDL

Table 1 (continued). Current studies on alcohol-induced metabolic perturbations. Shows the result of literature baseline analysis using PubMed

2. Objectives

The overall objective of this thesis was to extend the knowledge in the research area of alcohol biomarker discovery. Here, the novel high-throughput mass spectrometry platform was used on a big European human population cohort to perform a metabolomics study and gain new insights on alcohol-induced metabolic changes in human and their potential underlying biochemical mechanism. Following aims were addressed:

2.1 Identification of candidate alcohol biomarkers

Perform an analysis of the given human population data and metabolomics data to identify and characterize alcohol—induced metabolic signatures and select alcohol-related candidate biomarkers

2.2 Biochemical and functional interpretation

Identify potential underlying biochemical pathways of the alcohol-related metabolic candidate biomarker panel and elucidate potential underlying alcohol-related biochemical mechanisms.

2.3 Clinical utility

Investigate the utility of the mass-spectrometry-based setup in the context of biomarkers and technology for clinical diagnostics. Identify potential scenarios for the utility of metabolomics-based biomarkers. Go a step further beyond the “traditional” target discovery and data interpretation and discuss the capability of required key steps to bring the selected candidate biomarkers from bench-to-bedside.

3. Materials and Methods

3.1 Baseline research and literature reviews

For the literature review of alcohol-related metabolic perturbations PubMed (<https://www.ncbi.nlm.nih.gov> , (NCBI) 2010) was used to search available articles on studies related to metabolomics-related alcohol research between 1900 - 2015. PubMed search terms were “metabolomics” AND “alcohol”, “ethanol”, “metabolic profiling”, “metabolic” in Title/Abstract. Sources were selected through manual scanning of the Title/Abstracts for biomarker studies. Altogether 35 relevant sources could be identified and further analysed. For the final literature review 18 original research studies were included.

3.2 Metabolomics study design

For the targeted metabolomics setup a case-control study was conducted to compare differences of metabolite concentrations between different alcohol consumption groups.

3.3 Study populations

3.3.1 KORA F4 discovery data set

Cooperative Health Research in the Region of Augsburg (KORA) is a population-based research platform with subsequent follow-up studies in the fields of epidemiology and health care research (Holle, Happich et al. 2005, Lowel, Doring et al. 2005, Wichmann, Gieger et al. 2005). The KORA F4 study is the follow-up of KORA-Survey 4 (S4, 1999/2001) conducted in 2006/2008. 3080 individuals participated in the follow-up study. For 3061 individuals, metabolic data was available (Wang-Sattler, Yu et al. 2012, Yu, Zhai et al. 2012). From 3061 individuals, 1144 males and 946 females aged 32-81 were selected for further analysis after

application of the following exclusion criteria: Non-fasting at examination, diabetic, alcohol abstainer, missing data or outliers (i.e. extreme low or high values) in metabolite concentration data (**Figure 5**) (see also section 3.6.1 for outlier detection calculation and section 3.6.2 missing values). Study participants were categorized according to daily alcohol intake as light drinkers (LD) (females <20g/day and males <40g/day) and moderate-to-heavy drinkers (MHD) (females \geq 20g/day and males \geq 40g/day).

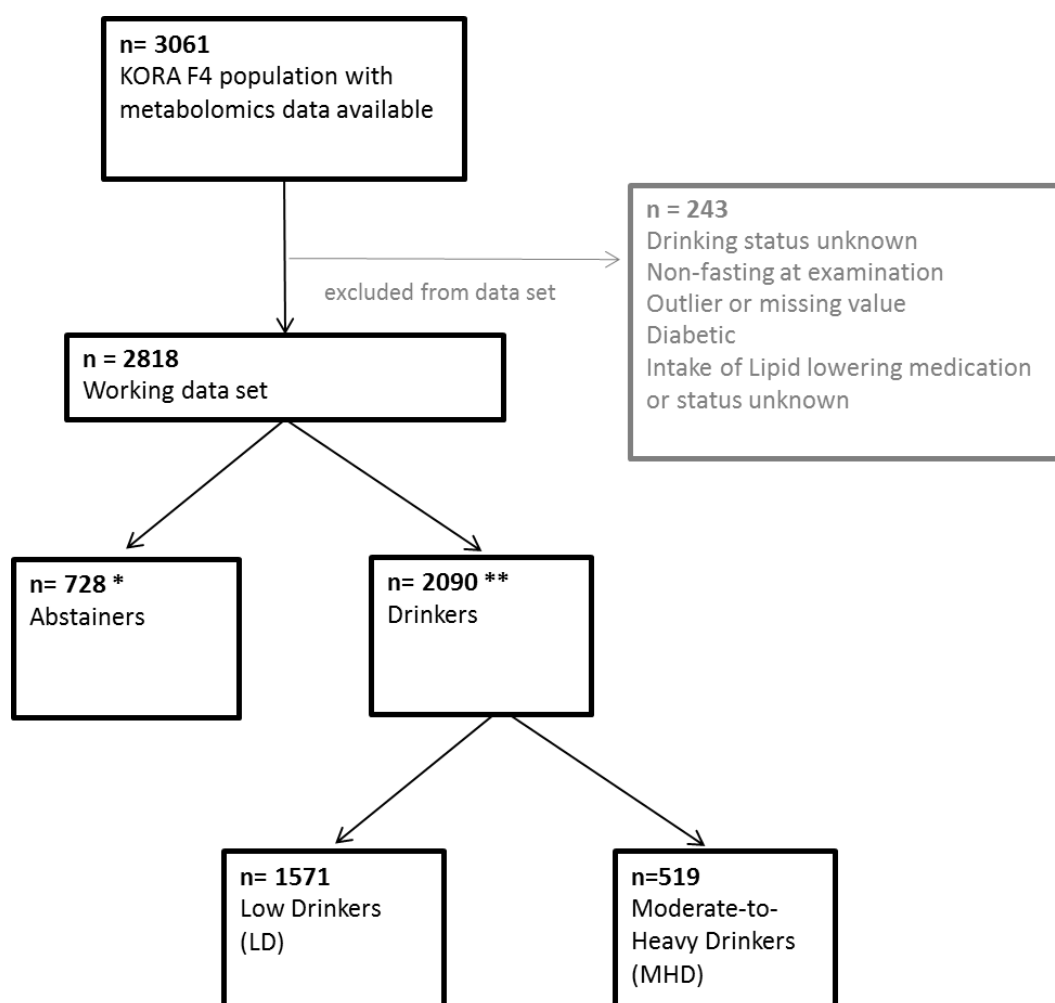


Figure 5. Selection process of study participants. Flowchart displays selection steps of data sets used for the subsequent analyses. (**) Data set used in the main investigation of alcohol-induced metabolic changes. Drinkers encompass LD and MHD. (*) Data set used for sub-analyses

3.3.2 KORA F3 replication data set

The KORA F3 study is a follow-up of the KORA-Survey 3 (S3, examined 1994/95), conducted in 2004/05. The KORA F3 cohort is a ten years follow-up survey of the KORA S3 survey. 2974 individuals participated in the follow-up. From 2974 individuals, 377 individuals had metabolic data available 154 males and 107 females aged 55-84 were selected for further analysis after the application of KORA F4 exclusion criteria. KORA F4 and KORA F3 are two independent cohorts and do not contain common participants and were conducted at different time points (Mittelstrass, Ried et al. 2011, Yu, Kastenmuller et al. 2011)

3.3.3 TwinsUK replication data set

The UK Adult Twin Registry (TwinsUK) is a UK-wide twin registry sample of 11 000 adults founded in 1993 with the aim to explore the genetic epidemiology of common adult diseases (Moayyeri, Hammond et al. 2012). 629 individuals aged 23-73 were selected for analysis after the application of KORA F4 exclusion criteria. For 277 probands HDL data was available.

3.3.4 Data collection

In KORA studies, the information on socio-demographic variables, smoking habits, physical activity, medication use, alcohol consumption was gathered by trained medical staff during a standardized interview (Ruckert, Heier et al. 2011)

3.3.5 Assessment of alcohol consumption

Assessment of alcohol intake (in grams per day) was based on weekday and weekend consumption of beer, wine and spirits and study participants provided information on their smoking behavior (never, past, current). Furthermore, participants underwent an extensive standardized medical examination including the

collection of blood samples. All measurement procedures were described in detail elsewhere (Ruckert, Heier et al. 2011).

3.3.6 Ethics Statement

Written informed consent has been given by each KORA and TwinsUK participant. The KORA studies, including the protocols for subject recruitment and assessment and the informed consent for participants, were reviewed and approved by the local ethical committee (Bayerische Landesärztekammer). For the TwinsUK study, ethics approval was received from the St Thomas' Hospital Ethics Committee.

3.4 Blood sampling

KORA blood samples for metabolic analysis were collected together with medical examinations as described in (Illig, Gieger et al. 2010, Jourdan, Petersen et al. 2012). KORA F4 blood samples were drawn into serum tubes in the morning between 8:00 and 10:30am after overnight fasting. Tubes were gently inverted twice, followed by 30 minutes resting at room temperature to obtain complete coagulation. For serum collection, centrifugation of blood was performed for 10 min (2 750 g, 15 °C). Serum was frozen at -80 °C until execution of metabolic analyses. TwinsUK blood samples were taken after at least 6 hours of overnight fasting. The samples were immediately inverted three times, followed by 40 min resting at 4°C to obtain complete coagulation. The samples were then centrifuged for 10 min at 2 000g. Serum was removed from the centrifuged brown-topped tubes as the top, yellow, translucent layer of liquid. Four aliquots of 1.5 ml were placed into skirted micro centrifuge tubes and then stored in a -45 °C freezer until sampling.

3.5 Metabolite measurements

Metabolomic analysis was performed on 3061 subjects from the KORA F4 study, 377 subjects from the KORA F3 study and 629 TwinsUK study. The targeted metabolomic approach was based on ESI-FIA-MS/MS measurements by AbsoluteIDQ™ p150 assay (BIOCRATES Life Sciences AG, Innsbruck, Austria). The method of AbsoluteIDQ™ p150 assay has been proven to be in conformance with FDA-Guideline "Guidance for Industry - Bioanalytical Method Validation (May 2001)" (Health, Human et al. 2001), which implies proof of reproducibility within a given error range. The assay procedures of the AbsoluteIDQ™ p150 kit as well as the metabolite nomenclature have been described in detail previously (Illig, Gieger et al. 2010, Römisch-Margl, Prehn et al. 2011, Wang-Sattler, Yu et al. 2012). Data evaluation for quantification of metabolite concentrations and quality assessment is performed with the MetIQ™ software package, which is an integral part of the AbsoluteIDQ™ kit. Internal standards serve as reference for the calculation of metabolite concentrations. To ensure data quality, each metabolite had to meet the three criteria described previously (Mittelstrass, Ried et al. 2011): (1) average value of the coefficient of variance (CV) for the metabolite in the three QCs should be smaller than 25%; (2) 90% of all measured sample concentrations for the metabolite should be above the limit of detection (LOD); and (3) the correlation coefficient between two duplicate measurements of the metabolite in 144 re-measured samples should be above 0.5. In total, 131 metabolites passed the three quality controls, and the final metabolomics dataset contained the sum of hexoses (H1), 14 amino acids, 24 acylcarnitines, 13 sphingomyelins, 34 diacylphosphatidylcholines, 37 acyl-alkyl-phosphatidylcholines, and 8 lysophosphatidylcholines.

3.6 Data pre-processing

Data pre-processing is essential for data analysis and interpretation of large metabolomics data sets. Metabolomics experiments generate a wide variety of data (for example concentrations of metabolites) for each individual sample. These data have to be streamlined and simplified for further bio statistical analyses. The first main goal is to check if the data is valid and present them in a meaningful way (Hocher and Adamski 2017). Data pre-processing and following statistical data analyses were performed with the open source software R (version 3.3.1).

3.6.1 Outlier detection

To identify outliers in the metabolomics data, concentrations obtained for the 131 metabolites were first scaled to zero mean and unity standard deviation and were projected onto the unit sphere, and Mahalanobis distances for each individual were then calculated using the robust principal components algorithm (Filzmoser, Maronna et al. 2008). Calculations were done separately for males and females. For each group, the mean Mahalanobis distance plus three times variance were defined as the cut-off. This algorithm utilizes simple properties of principal components to identify outliers in the transformed space, leading to significant computational advantages for high dimensional data (Filzmoser, Maronna et al. 2008). This approach requires considerably less computational time than existing methods for outlier detection, and is suitable for use on very large data sets (Filzmoser, Maronna et al. 2008).

3.6.2 Missing value imputation

Missing values in metabolomics data sets can originate from both technical and biological reasons and can lead to reduction of statistical power and can introduce bias in epidemiological studies. Here, the missing values were imputed using the R package “mice”(van Buuren and Groothuis-Oudshoorn 2012). The R package “mice”

imputes incomplete multivariate data by chained equations. The approach of multivariate imputation by chained equations was chosen because it handles missing values in more than one variable and takes into consideration correlations between metabolites or observations.

3.6.3 Data normalization

The normalization procedure aims to reduce the impact of very large feature values (that typically occur in metabolomics experiments with many different metabolites) and to make all features (i.e. metabolite concentrations) more comparable or normally distributed. Here, the metabolite concentrations were normalized by log-transformation for all subsequent analysis steps.

3.6.4 Data distribution check

When comparing experimental data sets (for example control versus patient group) the sets must display enough diversity in the data for each parameter studied. Shapiro-Wilk test (Royston 1982) normality test was applied on single metabolites to check for normal distribution of metabolites in the study population in order to choose proper follow-up tests. Mann-Whitney test (Bauer 1972) was applied for the comparison of two variables not satisfying normal distribution. Fisher's exact test (Agresti 2002) was applied for comparing binomial proportions.

3.7 Statistical data analysis

3.7.1 Logistics regression analysis

The logistic regression model predicts the probabilities of a sample being a member of either of two groups for a set of metabolite concentrations. The response variable of each sample for the logistic regression model, is binary, corresponding to the two groups (i.e. Abstainer, moderate-to-heavy (MHD), drinker or low-drinker (LD)).

Logistic regression (Hastie and Pregibon 1992) (adjusted for age, BMI, smoking, HDL and triglycerides) was applied on each of the 131 metabolites to investigate the association between the metabolite concentration and the alcohol consumption groups of Abstainer, LD and MHD respectively. To handle false discovery from multiple comparisons, the cut-off point for significance was calculated according to the Bonferroni correction. P-values were corrected according to the Bonferroni correction, at a level of $3.8E-4$ (for a total use of 131 metabolites at the 5% level). Odds ratios (ORs) for single metabolites were calculated between two groups. The concentration of each metabolite was scaled to have a mean of zero and an s.d. (i.e. standard deviation) of one; thus, all reported, OR values correspond to the change per s.d. of metabolite concentration. Various factors were added to the logistic regression analysis as covariates to identify the best model.

3.7.2 Random forest selection

To further select candidate biomarkers, two additional methods were applied (Wang-Sattler, Yu et al. 2012, Yu, Zhai et al. 2012): the random forest selection (Breiman 2001) and the stepwise selection, which assess the metabolites as a group. Between two groups, the random forest was first used to select the metabolites among the 30 highest ranking variables of importance score, allowing the best separation of the individuals from different groups. Age, BMI, smoking, HDL and triglycerides were also included in this method with all the metabolites

3.7.3 Stepwise selection

The metabolites were further selected using stepwise selection on the logistic regression model (**Figure 6**). Metabolites with significantly different concentrations between the compared groups in logistic regression, and which were also selected using random forest, were used in this model along with all the co-variates. Both

Random forest and stepwise selection can be seen as complementary to each other.

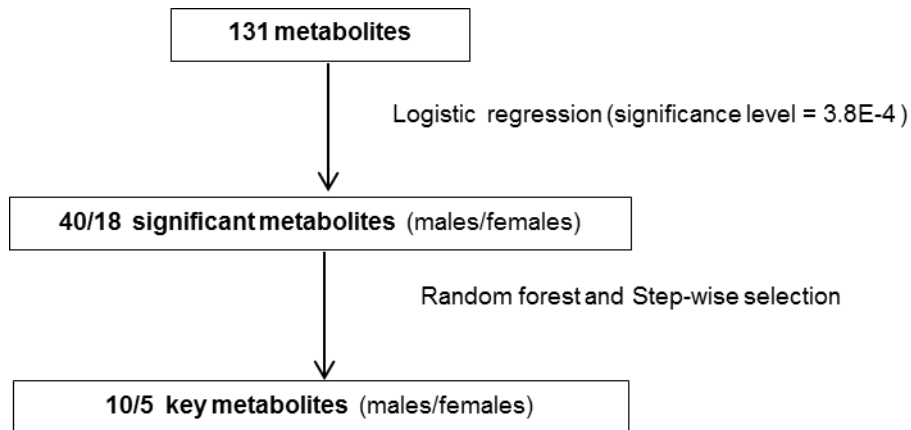


Figure 6. Selection path of key metabolites

3.7.4 Akaike's Information Criterion (AIC)

Akaike's Information Criterion (AIC) was used to evaluate the performance of these subsets of metabolites used in the models. The model with minimal AIC was chosen.

3.7.5 Correlation analysis

A correlation exists between two variables when one of them is related to the other. Pearson's correlation coefficient (r) measures the strength of the linear relationship between the paired x - and y -quantitative values in a sample if both x and y follows a normal distribution. For scenarios where the distribution is not normal, Spearman correlation (ρ) can be calculated instead (www.statisticssolutions.com). For the correlation analysis between metabolite concentrations and laboratory parameters (HDL and triglycerides) the Spearman correlation was chosen.

3.7.6 Area under the receiver-operating characteristic (AUROC)

The area under the receiver-operating-characteristic curves (AUROC) was used to evaluate the logistic regression models. The area under the receiver operating curve is a measure of classification model performance (i.e. AUROC close to 1 indicates a

successful classification model).

3.7.7 Clustering and Heat maps

Heat maps were used to illustrate the trends of metabolite concentrations with increasing alcohol consumption. Heat map is a visualization technique which converts numerical table into a corresponding 2D color map to provide an intuitive overview of the data values. Heat map are often used with hierarchical clustering to reorganize the rows/columns of the data with dendograms plotted to the corresponding side. Alcohol consumption data was split into alcohol consumption categories increasing by 5g/day. A matrix of mean metabolite concentrations was calculated for each alcohol consumption category for significant male/female-specific metabolites from logistic regression. In the same procedure step hierarchical clustering with Euclidean distance was applied on the metabolite concentration matrix to generate a hierarchical dendogram clustering metabolites with similar mean metabolite concentrations.

3.7.8 Meta-analysis

For the meta-analysis of the combined KORA F4 and KORA F3 studies a fixed effect model was used.

3.7.9 Smoother plots

Smoother plots were drawn for each metabolite of the set of metabolites with the R function 'qplot' (package 'ggplot2') using the options geom=smooth, method=loess, span=0.5 producing smoother plots with locally weighted regression (locally weighted scatterplot smoothing (loess)) applying a smoothing span of 0.5, which results in medium smoothing. loess computes outlier robust locally weighted regression fitted values by fitting local polynomials using the weights and results in the loess curve as

shown in our smoother plots. For better visualization the plots were truncated to observations between first and 99th percentile.

3.8 Metabolite biochemical and functional Interpretation

3.8.1 Metabolite biochemical interpretation

Biochemical interpretation of metabolomics data means in the context of the PhD project the description and characterization of the alcohol-related metabolites found by statistics in a biological context. The Biocrates p150 kit generally includes a product manual listing metabolite aliases/names as (e.g. PC aa C32:1), however in the product manual not all metabolite names have a respective unique identifier and are up-to-date from the metabolomics database as HMDB and KEGG etc. For example, Biocrates uses same KEGG ID for two different metabolites (i.e. SM(OH)C16:1 and SM(OH)C22:2). This is why a more comprehensive manual query for unique identifiers for each metabolite was conducted using a bigger range of metabolomics sources in PubChem (Kim, Thiessen et al. 2016) MetaboAnalyst 2.0 (Xia, Mandal et al. 2012), NCBI (Jenuith 2000) HMDB 3.0 (Wishart, Jewison et al. 2013).

3.8.2 Metabolic Pathway and Functional Enrichment analysis

Metabolic pathway analysis was performed using MetaboAnalyst 2.0 (Xia, Mandal et al. 2012). The metabolic pathway analysis module in MetaboAnalyst 2.0 combines results from pathway enrichment analysis with the pathway topology analysis to help identify the most relevant pathways involved in the conditions or phenotype studied (i.e. affected by moderate-to heavy alcohol consumption). It uses metabolic pathways from the KEGG (Ogata, Goto et al. 1999) library.

3.8.3 Literature research on alcohol-related metabolites and associated disease states and phenotypes

For the identified alcohol-related 10/5 key metabolites in males/females a systematic manual literature analysis was conducted in PubMed for reported associations between the metabolites and other disease states and phenotypes.

4. Results

4.1 Baseline characteristics of the discovery and replication data sets (KORAF4, KORAF3 and TwinsUK)

4.1.1 Drinkers

The main investigation of alcohol-induced metabolic alterations was focused on the drinkers group. The drinkers group made approximately 74%/ 70%/74% of the total population in KORAF4/KORAF3/TwinsUK. A strictly sex-separated analysis was conducted based on prior results from KORA F4, which showed strong metabolomic differences between men and women (Mittelstrass, Ried et al. 2011). There were more male drinkers than female drinker across KORA F4 and KORA F3 studies. For both sexes, the probands were classified into two groups according to average daily alcohol consumption of low drinkers (LD) and moderate-to-heavy drinkers (MHD) based on the common measures of the WHO International Guide for monitoring alcohol consumption and related harm (WHO 2000) and compared MHD with LD **(Table 2)**. Across KORA F4 and KORA F3 studies and across both sexes there were more low drinkers than moderate-to-heavy drinkers. In general, age and BMI were comparable between MHD and LD. A significantly lower age could be observed in MHD of KORA F3 males and TwinsUK participants (p-value 1.3E-02 and 1.6E-02, respectively). BMI was significantly increased in MHD in male KORA F4 participants (p-value 3.3E-03). The proportion of smokers was significantly higher in MHD in

KORA F4 male and TwinsUK female populations (p-values 1.0E-04 and 1.3E-02, respectively). In all three studies, there was a significant increase in HDL in MHD compared to LD (p-values 7.1E-12 – 1.3E-02). Except in KORA F3, the mean HDL was increased but p-value was not significant. Significant increase of mean triglyceride concentration could be observed in KORA F4 male MHD only (p-value 3.4E-02).

4.1.2 Abstainers

The Abstainer group was included in the sub-analyses. Altogether the abstainer group was approximately 26%/30%/26% of the total KORA F4/KORAF3/TwinsUK population (**Table 3**). There were more female abstainers compared to male abstainers in KORA F4 and KORA F3. Data was not available on the granularity level to investigate abstainer subgroups (i.e. duration of alcohol abstinence).

	Discovery KORA F4 (n=2090)						Replication KORA F3 (n=261)						Replication TwinsUK (n=277)		
	men (n=1144)			women (n=946)			men (n=154)			women (n=107)			women (n=277) ^c		
	LD ^a	MHD ^b	P-value	LD	MHD	P-value	LD	MHD	P-value	LD	MHD	P-value	LD	MHD	P-value
participants n(%)	859(75)	285(25)	--	712(75)	234(25)	--	106(69)	48(31)	--	84(78)	23(22)	--	250 (90)	27(10)	--
age(years) mean±SD	56.2±13.7	57.1±11.6	2.8E-01 ^d	54.5±13.2	55.7±11.5	1.8E-01 ^d	67.9±7.6	64.5±6.8	1.3E-02^d	65.5±7.4	64.6±7.7	5.3E-01 ^d	48.4±10.2	43.5±9.0	1.6E-02^d
BMI (kg/m ²) mean±SD	27.6±3.8	28.4±4.2	3.4E-03^d	26.6±4.9	26.3±4.7	3.6E-01 ^d	28.1±3.1	29.1±3.6	1.2E-01 ^d	27.6±3.7	27.5±4.0	9.5E-01 ^d	25.1±4.6	24.7±4.1	8.1E-01 ^d
smoker n(%)	153(18)	66(23)	1.0E-04^e	103(14)	44(19)	1.2E-01 ^e	8(7)	8(17)	9.5E-02 ^e	4(5)	3(13)	1.7E-01 ^e	67(27)	14(52)	1.3E-02^e
HDL(mg/dl) mean±SD	49.9±11.3	55.8±13.5	7.1E-12^d	62.7±13.6	66.0±15.0	4.4E-03^d	49.2±13.2	59.3±15.3	1.5E-04^d	65.7±16.7	71.10±16.2	1.4E-01 ^d	58.9±14.5	66.2±13.9	1.3E-02^d
triglyceride (mg/dl) mean±SD	134.5±82.0	154.0±104.9	3.4E-02^e	103.9±71.4	100.9±53.6	9.7E-01 ^e	188.0±106.7	192.3±130.2	9.6E-01 ^e	156.2±74.1	146.4±85.0	3.0E-01 ^e	99.6±47.3	94.1±71.4	1.1E-01 ^d

Table 2. Drinkers - study population characteristics of KORA F4 discovery, KORA F3 and TwinsUK replication data sets.

Drinking status defined as ^aalcohol consumption < 40g/day in males, < 20 g/day in females, ^balcohol consumption ≥ 40g/day in males, ≥ 20 g/day in females; Abbreviations: BMI=body mass index; HDL= high density lipoprotein, SD=standard deviation, LD=light drinkers, MHD=moderate-to-heavy drinkers, ^c629 participants were available for replication analysis, for 277 participants HDL-C and triglyceride data was available, ^dMann-Whitney test (two-sided), ^eFisher's exact test (two-sided), significance level α = 0.05. Significant p-values are in **bold**.

	Discovery KORA F4 (n=728)		Replication KORA F3 (n= 109)		Replication TwinsUK (n=99)
	men (n=233)	women (n=495)	men (n=37)	women (n=72)	women (n=99)
participants n(%)	233 (32)	495 (68)	37 (33.94)	72 (66.055)	99 (100%)
age(years) mean±SD	53.7 ± 14.0	54.5 ±13.2	65.4 ± 5.8	66.4 ± 7.7	51.3 ± 9.5
BMI (kg/m ²) mean±SD	27.7 ± 4.6	27.7 ± 5.4	28.0 ± 3.8	29.9 ± 4.8	25.89 ± 5.0
smoker n(%)	61 (26.180)	86 (17.37)	6 (16.21622)	14 (19.444)	28 (28.2823)
HDL(mg/dl) mean±SD	45.8 ± 11.1	58.4 ± 14.1	48.4 ± 10.3	59.6 ± 16.9	53.1 ± 13.2
triglyceride (mg/dl) mean±SD	141.5 ± 134.2	105.4 ± 55.3	189.7 ± 142.0	173.5 ± 107.3	138.8 ± 108.7

Table 3. Abstainers - study population characteristics of KORA F4 discovery, KORA F3 and TwinsUK replication data sets. Defined as alcohol consumption = 0g/day (males and females); Abbreviations: BMI= body mass index; HDL= high density lipoprotein; SD= standard deviation

4.2 Global metabolic differences between Abstainers, Low-drinkers and Moderate-to-heavy drinkers

Logistic regression analysis (adjusted for age, BMI, smoking, HDL and triglycerides) identified 40 significant metabolites in males and 18 metabolites in females that significantly differed (p -value $< 3.8E-4$) in concentration between only MHD and LD in the KORA F4 study. No significant metabolic changes were found between LD and Abstainers (Figure 7) (Figure 8), see (Table 4) (Table 5) for detailed p -values and direction).

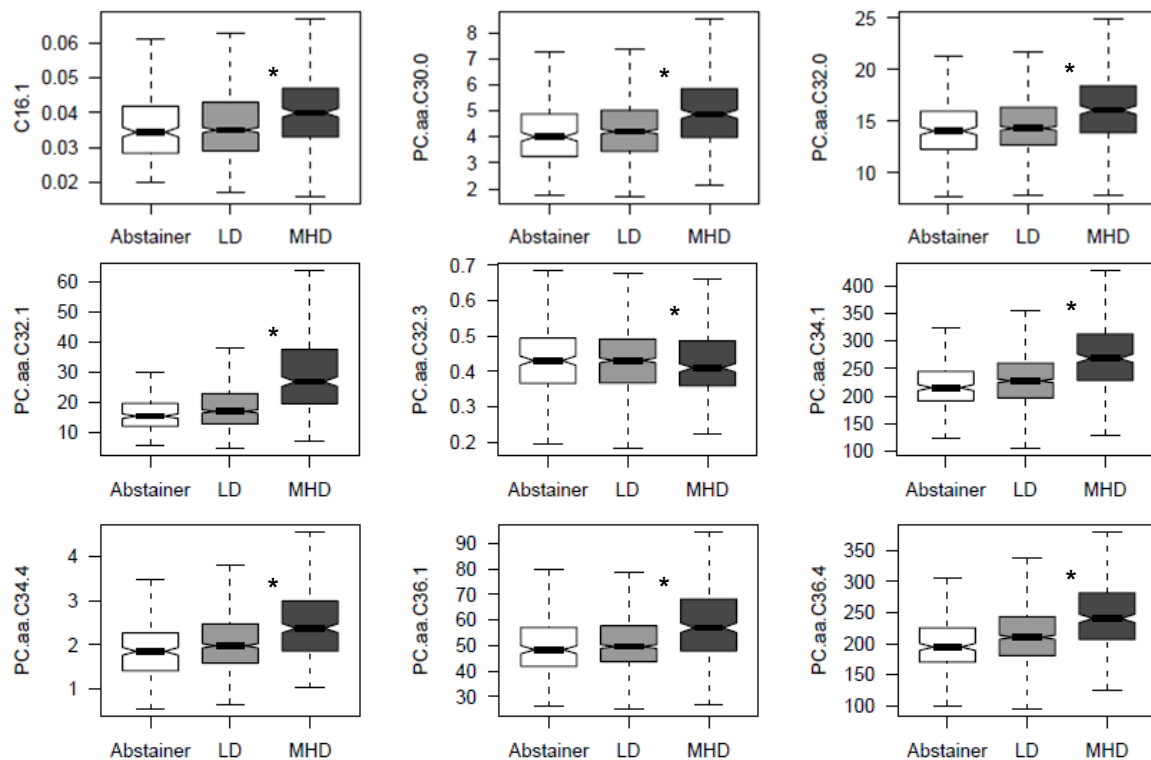


Figure 7. Global analysis in males shows significant metabolic changes only between moderate-to-heavy drinkers (MHD) and low-drinkers (LD). Boxplot representation of 40 significant metabolites identified in KORA F4 by logistic regression. Displayed are mean absolute metabolite concentrations of all three alcohol consumption groups (Abstainer, LD and MHD). Significant changes between two groups are marked with an asterix (*).

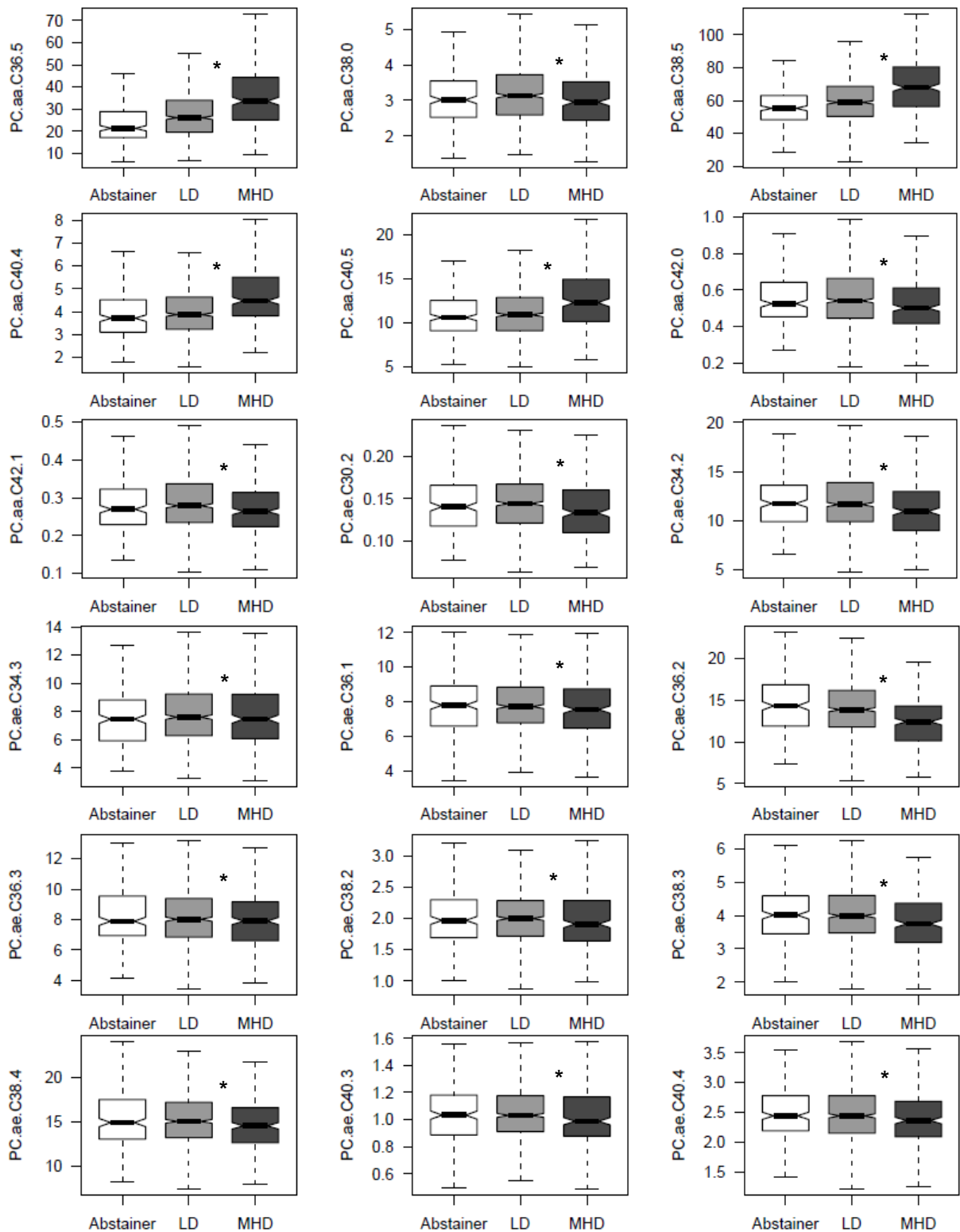


Figure 7(continued). Global analysis in males shows significant metabolic changes only between moderate-to-heavy drinkers (MHD) and low-drinkers (LD). Boxplot representation of 40 significant metabolites identified in KORA F4 by logistic regression. Displayed are mean absolute metabolite concentrations of all three alcohol consumption groups (Abstainer, LD and MHD). Significant changes between two groups are marked with an asterisk (*).

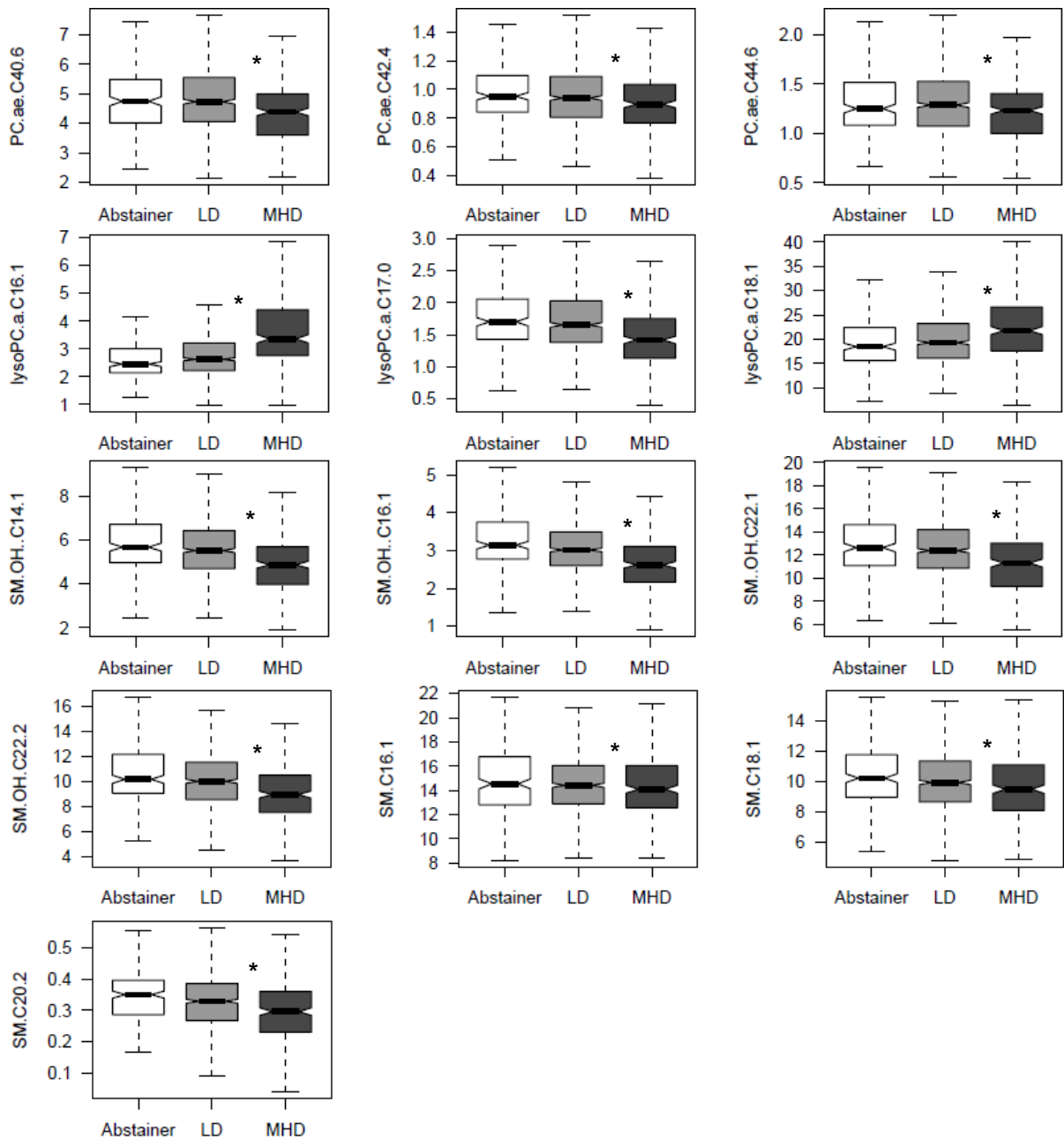


Figure 7 (continued). Global analysis in males shows significant metabolic changes only between moderate-to-heavy drinkers (MHD) and low-drinkers (LD). Boxplot representation of 40 significant metabolites identified in KORA F4 by logistic regression. Displayed are mean absolute metabolite concentrations of all three alcohol consumption groups (Abstainer, LD and MHD). Significant changes between two groups are marked with an asterix (*).

Metabolite	Abstainer	LD	MHD	LD vs Abstainer		MHD vs LD	
	Mean ± SD (µM)	Mean ± SD (µM)	Mean ± SD (µM)	Odds ratio (95% CI)	P-value	Odds ratio (95% CI)	P-value
C16:1	0.0362 ± 0.0104	0.0372±0.0109	0.0419±0.0128	0.99	9.33E-01	1.34	1.43E-04
PC aa C30:0	4.23±1.34	4.35±1.29	5.06±1.55	1.01	9.39E-01	1.37	4.19E-05
PC aa C32:0	14.40±3.20	14.67±2.82	16.44±3.59	1.04	8.26E-01	1.40	3.21E-05
PC aa C32:1	17.06±7.75	19.27±9.16	31.04±16.87	1.29	2.64E-01	2.34	1.15E-18
PC aa C32:3	0.4411±0.1054	0.4343±0.0918	0.4242±0.0955	1.16	3.75E-01	0.68	3.20E-06
PC aa C34:1	223.33±49.97	232.38±48.19	276.21±67.34	1.08	6.74E-01	1.96	9.25E-13
PC aa C34:4	1.88±0.64	2.08±0.72	2.49±0.83	1.67	1.05E-02	1.41	1.80E-05
PC aa C36:1	51.15±14.11	51.35±11.89	59.70±16.38	0.80	1.80E-01	1.48	6.63E-06
PC aa C36:4	199.44±46.24	215.33±49.60	245.90±54.86	1.61	1.62E-02	1.49	2.10E-06
PC aa C36:5	24.35±11.46	28.69±12.93	36.59±17.55	2.08	5.75E-03	1.46	3.98E-07
PC aa C38:0	3.12±0.84	3.23±0.91	3.03±0.81	1.00	9.94E-01	0.68	4.48E-06
PC aa C38:5	56.14±12.69	60.34±14.04	69.93±18.12	1.73	1.36E-02	1.55	2.10E-07
PC aa C40:4	3.98±1.17	4.04±1.17	4.85±1.59	1.30	1.55E-01	1.57	1.25E-07
PC aa C40:5	11.02±2.99	11.21±3.00	13.00±4.16	1.37	1.10E-01	1.45	1.86E-05
PC aa C42:0	0.56±0.16	0.57±0.18	0.52±0.15	0.89	5.24E-01	0.67	5.34E-06
PC aa C42:1	0.28±0.08	0.29±0.08	0.27±0.08	1.05	7.59E-01	0.73	2.18E-04
PC ae C30:2	0.65±0.15	0.15±0.04	0.14±0.04	1.08	6.65E-01	0.73	7.56E-05
PC ae C34:2	12.07±3.06	12.01±3.08	11.13±2.83	0.86	3.77E-01	0.56	6.04E-11
PC ae C34:3	7.73±2.31	7.93±2.28	7.73±2.22	1.17	4.18E-01	0.65	2.95E-06
PC ae C36:1	7.85±1.87	7.85±1.65	7.69±1.77	0.87	4.43E-01	0.73	9.89E-05

Table 4. Result of logistic regression model analysis in KORA F4 in males Odds ratios (OR) and p-values in two pairwise comparisons of alcohol consumption groups. Shows 40 significant metabolites in males that were found significantly different in concentration between MHD and LD in the KORA F4 discovery sample. Odds ratios (OR) reflect the pairwise comparison of metabolite concentrations between LD and Abstainers, MHD and LD. Out of the initial set of 131 metabolites, only 40 significant metabolites are shown. Significant P-values are highlighted in bold. No significant differences were found comparing LD with Abstainers

Metabolite	Abstainer	LD	MHD	LD vs Abstainer		MHD vs LD	
	Mean ± SD (µM)	Mean ± SD (µM)	Mean ± SD (µM)	Odds ratio (95% CI)	P-value	Odds ratio (95% CI)	P-value
PC ae C36:2	14.67±3.65	14.12±3.31	12.44±3.05	0.80	2.06E-01	0.46	1.97E-17
PC ae C36:3	8.25±1.95	8.20±1.97	7.96±1.85	0.83	2.66E-01	0.67	2.53E-06
PC ae C38:2	2.04±0.55	2.02±0.45	1.97±0.44	0.88	4.61E-01	0.68	3.04E-06
PC ae C38:3	4.10±0.95	4.04±0.84	3.82±0.86	0.92	6.47E-01	0.65	5.10E-08
PC ae C38:4	15.52±3.28	15.29±3.09	14.65±2.92	0.86	3.38E-01	0.71	1.61E-05
PC ae C40:3	1.06±0.26	1.05±0.21	1.02±0.22	0.89	5.34E-01	0.71	1.85E-05
PC ae C40:4	2.53±0.52	2.49±0.50	2.39±0.45	0.77	1.31E-01	0.72	6.98E-05
PC ae C40:6	4.83±1.14	4.89±1.18	4.42±1.08	0.99	9.56E-01	0.58	4.30E-10
PC ae C42:4	0.99±0.24	0.96±0.24	0.91±0.22	0.63	8.56E-03	0.69	2.09E-05
PC ae C44:6	1.32±0.36	1.33±0.38	1.25±0.35	0.83	3.02E-01	0.73	2.60E-04
lysoPC a C16:1	2.62±0.76	2.80±0.87	3.72±1.48	1.50	6.00E-02	1.99	8.63E-16
lysoPC a C17:0	1.77±0.50	1.73±0.50	1.48±0.49	1.04	8.15E-01	0.58	5.49E-10
lysoPC a C18:1	19.32±4.87	20.03±5.46	22.73±7.38	0.96	8.33E-01	1.44	1.67E-05
SM(OH)C14:1	5.95±1.58	5.62±1.32	4.95±1.33	0.75	1.14E-01	0.56	1.47E-11
SM(OH)C16:1	3.27±0.79	3.06±0.68	2.68±0.73	0.67	2.44E-02	0.56	6.85E-12
SM(OH)C22:1	12.93±2.85	12.56±2.52	11.41±2.85	0.90	5.39E-01	0.62	3.45E-09
SM (OH) C22:2	10.55±2.36	10.15±2.13	9.10±2.31	0.83	3.05E-01	0.53	4.07E-13
SM C16:1	14.89±2.75	14.55±2.45	14.35±2.62	0.87	4.12E-01	0.70	1.43E-05
SM C18:1	10.50±2.24	10.07±2.02	9.65±2.17	0.80	1.29E-01	0.68	1.40E-06
SM C20:2	0.35±0.09	0.33±0.09	0.29±0.10	1.00	9.82E-01	0.66	1.51E-07

Table 4 (continued). Result of logistic regression model analysis in KORA F4 in males. Odds ratios (OR) and p-values in two pairwise comparisons of alcohol consumption groups. Shows 40 significant metabolites in males that were found significantly different in concentration between MHD and LD in the KORA F4 discovery sample. Odds ratios (OR) reflect the pairwise comparison of metabolite concentrations between LD and Abstainers, MHD and LD. Out of the initial set of 131 metabolites, only 40 significant metabolites are shown. Significant P-values are highlighted in bold. No significant differences were found comparing LD with Abstainers

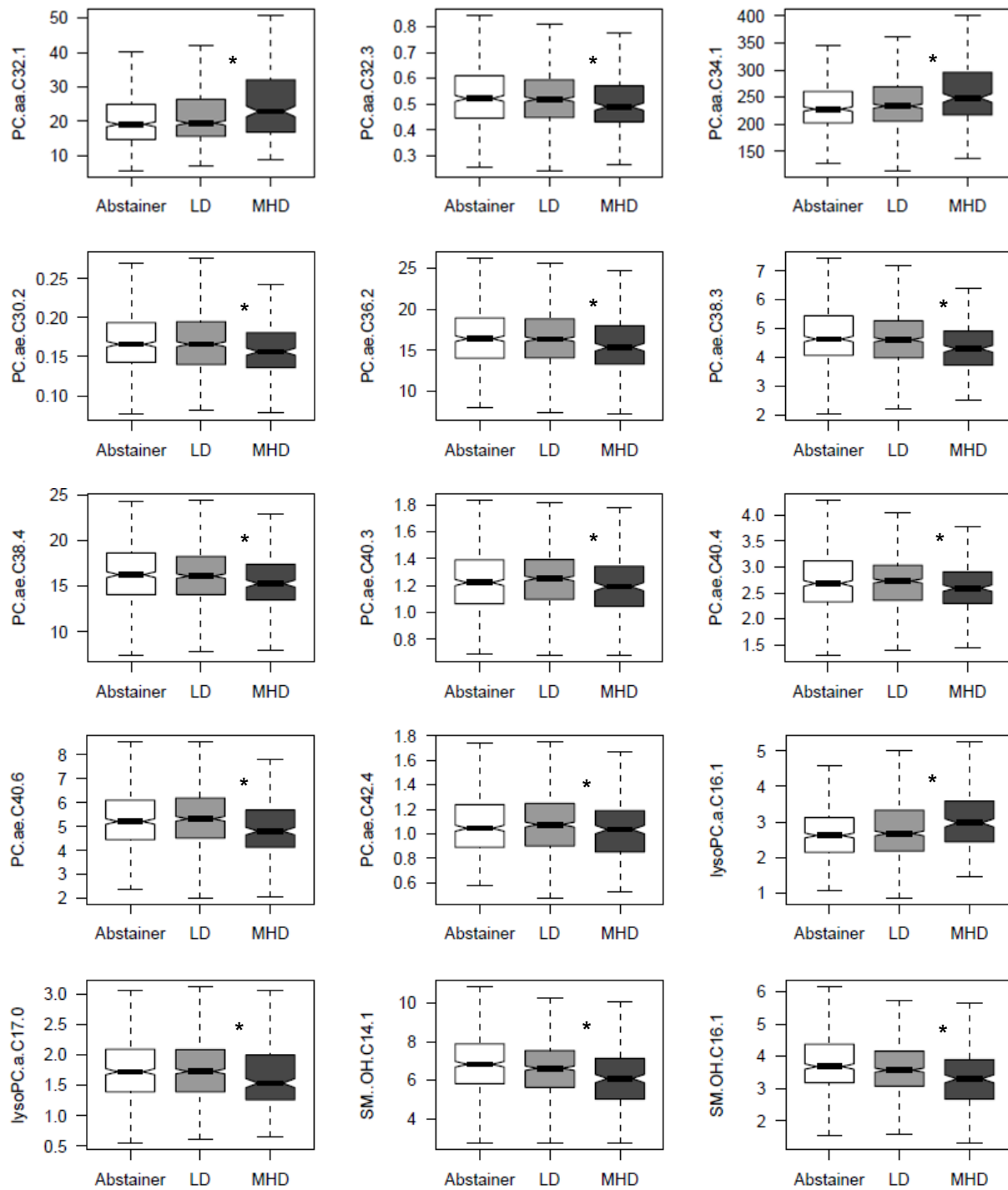


Figure 8. Global analysis in females shows significant metabolic change only between moderate-to-heavy drinkers (MHD) and low drinkers (LD). Boxplot representation of 18 significant metabolites identified in in KORA F4 females by logistic regression. Displayed are mean absolute metabolite concentrations of all three alcohol consumption groups (Abstainer, LD and MHD). Significant changes between two groups are marked with an asterisk (*).

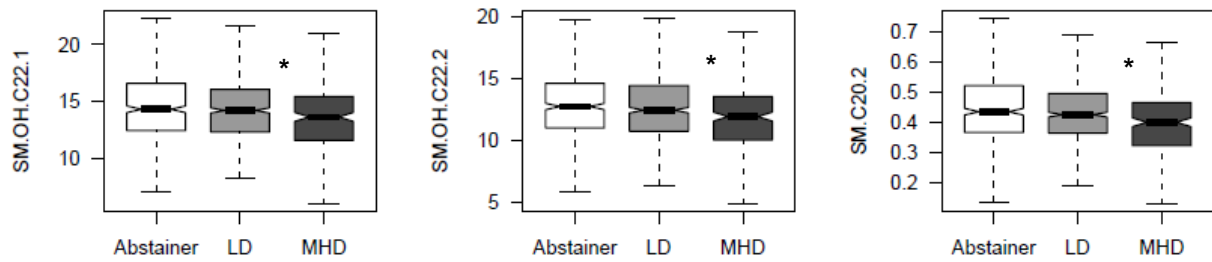


Figure 8 (continued). Global analysis in females shows significant metabolic changes only between moderate-to-heavy drinkers (MHD) and low drinkers (LD). Boxplot representation of 18 significant metabolites identified in in KORA F4 females by logistic regression. Displayed are mean absolute metabolite concentrations of all three alcohol consumption groups (Abstainer, LD and MHD). Significant changes between two groups are marked with an asterisk (*).

Metabolite	Abstainer	LD	MHD	LD vs Abstainer		MHD vs LD	
	Mean ± SD (µM)	Mean ± SD (µM)	Mean ± SD (µM)	Odds ratio (95% CI)	P-value	Odds ratio (95% CI)	P-value
PC aa C32:1	20.94±8.81	22.03±9.78	26.83±13.65	1.43	1.02E-01	1.56	2.54E-07
PC aa C32:3	0.5347± 0.1216	0.5275±0.1163	0.5084±0.1129	0.95	7.80E-01	0.72	2.56E-04
PC aa C34:1	233.62±47.30	241.12±52.86	259.38±56.36	1.28	2.32E-01	1.42	1.02E-04
PC ae C30:2	0.1694±0.0391	0.1703±0.0420	0.1594±0.0361	1.25	1.98E-01	0.70	4.04E-05
PC ae C36:2	16.77±3.80	16.67±3.77	15.61±3.58	0.90	5.38E-01	0.62	1.51E-07
PC ae C38:3	4.75±1.07	4.67±0.97	15.61±3.58	0.82	2.73E-01	0.68	6.62E-06
PC ae C38:4	16.45±3.39	16.30±3.09	15.50±3.08	0.72	6.58E-02	0.70	1.99E-05
PC ae C40:3	1.24±0.25	1.25±0.24	1.21±0.26	0.88	4.93E-01	0.73	2.69E-04
PC ae C40:4	2.74±0.55	2.74±0.51	2.62±0.51	0.76	1.13E-01	0.72	9.06E-05
PC ae C40:6	5.36±1.31	5.42±1.26	5.02±1.23	1.21	3.32E-01	0.63	2.36E-07
PC ae C42:4	1.08±0.25	1.09±0.26	1.04±0.25	0.85	3.40E-01	0.71	1.54E-04
lysoPC a C16:1	2.72±0.77	2.82±0.88	3.15±1.01	1.59	1.99E-02	1.36	4.67E-05
lysoPC a C17:0	1.76±0.50	1.76±0.50	1.63±0.49	1.21	2.68E-01	0.73	2.76E-04
SM(OH) C14:1	6.95±1.65	6.72±1.52	6.22±1.62	0.88	4.78E-01	0.62	8.98E-08
SM(OH)C16:1	3.79±0.88	3.65±0.81	3.36±0.87	0.83	2.94E-01	0.62	4.69E-08
SM(OH)C22:1	14.58±3.05	14.31±2.87	13.72±2.92	0.85	3.29E-01	0.74	3.20E-04
SM(OH)C22:2	12.91±2.70	12.72±2.64	11.97±2.67	0.83	3.44E-01	0.63	3.39E-07
SM C20:2	0.4476±0.1189	0.43±0.11	0.40±0.13	0.90	5.84E-01	0.72	9.71E-05

Table 5. Result of logistic regression model analysis in KORA F4 females. Odds ratios (OR) and p-values in two pairwise comparisons of alcohol consumption groups. Shows 18 significant metabolites in females that were found significantly different in concentration between MHD and LD in the KORA F4 discovery sample. Odds ratios (OR) reflect the pairwise comparison of metabolite concentrations between LD and Abstainers, MHD and LD. Out of the initial set of 131 metabolites, only 18 significant metabolites are shown. Significant P-values are highlighted in **bold**. No significant differences were found comparing LD with Abstainers.

4.3 Cluster analysis of significant global metabolic changes in Low-Drinkers and Moderate-to-Heavy drinkers

To illustrate the trend of metabolite levels with increasing alcohol consumption, heat maps were drawn based on normalized mean metabolite residuals for each of the 40/18 male/female metabolites. Hierarchical clustering with Euclidean distance was used in order to find similar metabolite groups. The final clusterogram (display of dendrogram and heat map) resulted in two main clusters C1 and C2 both in males and females (**Figure 9**). C1 consists of metabolites that increase in concentration with increasing alcohol consumption (high in MHD and low in LD). In contrast, C2 consist of metabolites that decrease in concentration with increasing alcohol consumption (low in MHD and high in LD). Diacylphosphatidylcholines (PC aa Cx:y)s, ether lipids (PC ae Cx:y)s, lysophosphatidylcholines (lysoPC a Cx:y)s, and sphingomyelins (SM)s occurred in both males and females. Only the acylcarnitine C16:1 occurred in males. All PC ae Cx:ys and SMs were decreased in MHD in males and females. PC aa Cx:ys were increased in MHD compared to LD in males and females (except PC aa C32:3 which was decreased in MHD in females. All lysoPC a Cx:ys were increased in MHD in males and females (except lysoPC a C17:0). For the completeness of analysis, the Abstainer group was also added to the cluster analysis. For the abstainer group, the same consistent clusters could be observed.

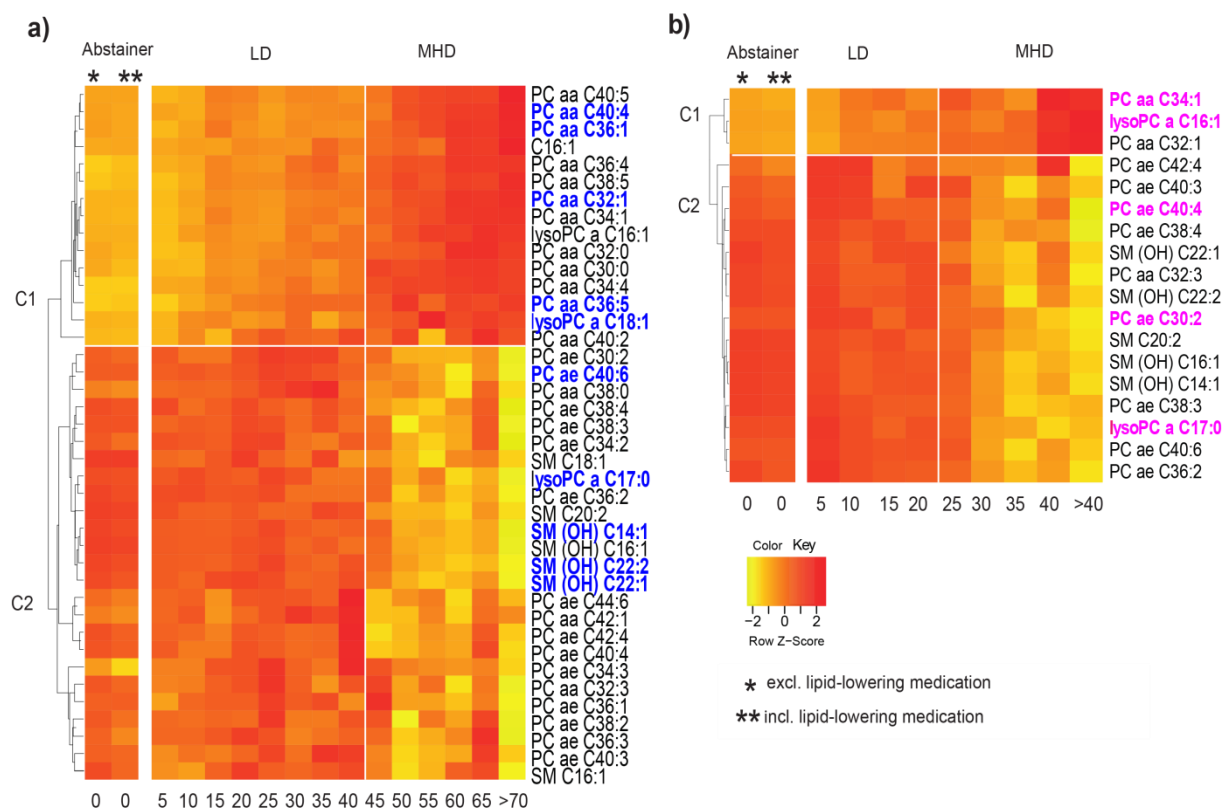


Figure 9. Alcohol-specific metabolite clusters and key metabolites. Clusterograms show metabolite concentrations in relation to alcohol consumption in light drinkers (LD) and moderate-to-heavy drinkers (MHD) of (a) 40 significant metabolites in males and (b) 18 significant metabolites females. The separate two-column clusterogram on the left shows the result of an additional sub-analysis for the abstainer group, displaying the effect of lipid-lowering medication (i.e. statins, fibrates, herbal-based lipid-lowering agents) on metabolite concentrations. Relative concentration of metabolites is represented by x-fold standard deviation from overall mean concentrations from groups of alcohol consumption of 5g per day. Horizontal axis displays alcohol concentration in g per day, while the vertical axis represents hierarchical clustering. Most specific 10/5 key metabolites separating MHD from LD in males/females are highlighted in blue/pink

4.4 Identification of alcohol-related key metabolites

4.4.1 Metabolite selection

Metabolomics data set is highly dimensional, with a broad number of features (peaks) ranging from few hundreds to few thousands. They represent a snapshot of global biochemical profiles of individual organisms. Most of these features are expected within normal variations, and only a few may be significantly associated with the condition or phenotype of interest. The identification of those “key” features is the first

step toward finding useful biomarkers or explaining the underlying biological process. The logistic regression analysis was based on each single metabolite, and some of these 40/18 male/female metabolites are expected to correlate with each other. To find more specific and independent metabolites that best separate MHD from LD as potential biomarkers for alcohol consumption, a further analysis applying random forest and stepwise selection method was used. Thus, ten metabolites in males (PC aa C32:1, PC aa C36:1, PC aa C36:5, PC aa C40:4, PC ae C40:6, lysoPC a C17:0, lysoPC a 18:1, SM (OH) C22:1, SM (OH) C22:2, SM (OH) C16:1) and five metabolites in females (PC aa C34:1, PC ae C30:2, PC ae C40:4, lysoPC a C16:1, lysoPC a 17:0) were further selected (**in Figure 9**: key metabolites are highlighted in pink for women and blue for men).

4.4.2 Model evaluation

To evaluate the model (i.e. combination of the ten/five male/female specific metabolites with covariates, in order to investigate how good does the logistic regression model adjusted for age, BMI, smoking, HDL and triglycerides distinguish between MHD and LD, area under the receiver-operating characteristic (AUROC or AUC) was calculated. The ROC curve analysis is generally considered the standard method for a performance assessment and is needed when developing new biomarkers and translating those biomarkers into clinical practice. The AUROC (or AUC) can be interpreted as the probability of how good the given logistic regression model will correctly classify i.e. (LD or MHD) a randomly chosen participant based on the alcohol-specific metabolite markers. The perfect model/classifier reaches an AUC=0.9-1.0, AUC=0.8-0.9 is “good”, AUC=0.7-0.8 is “fair”, AUC=0.6-0.7 is “poor”; AUC=0.5-0.6 is “poor”. In the current study the AUC value in males was 0.812 and in females 0.679 indicating that the alcohol-specific metabolite profile of ten/five

markers in men/females is a good-to-fair predictor for chronic moderate-to-heavy alcohol consumption in males/females (**Figure 10**).

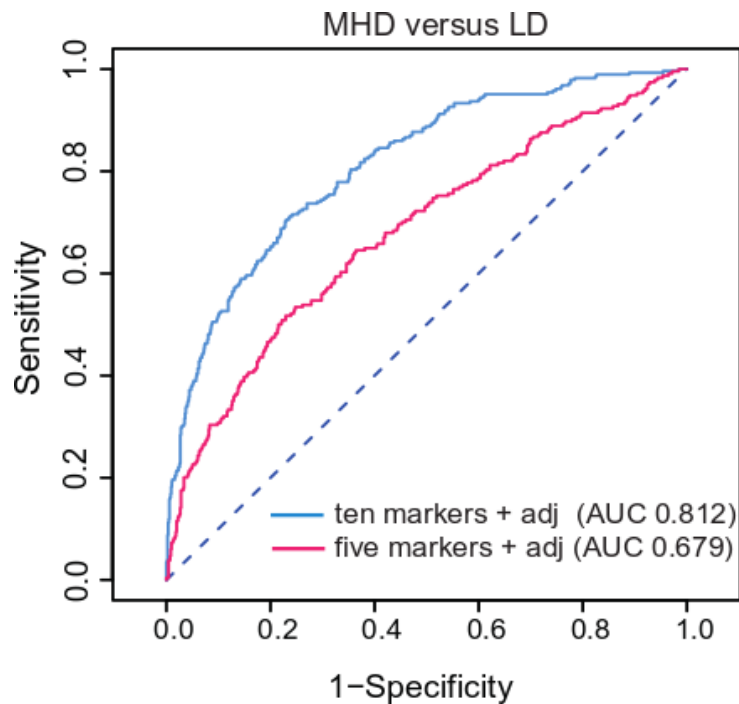


Figure 10. Alcohol-specific key metabolite panel gives a good-to-fair prediction for alcohol consumption. Graphics shows receiver operating characteristic (ROC) curves for the set of most significant key 10/5 metabolites in males (blue curve) (i.e. PC aa C32:1, PC aa C36:1, PC aa C36:5, PC aa C40:4, PC ae C40:6, lysoPC a C17:0, lysoPC a 18:1, SM(OH)C22:2, SM(OH)C22:1, SM(OH)C16:1) and females (pink curve) (i.e. PC aa C34:1, PC ae C30:2, PC ae C40:4, lysoPC a C16:1, lysoPC a 17:0). ROC curve displayed as dotted line/crossed line represent metabolite marker profile performance in males/females. The area under the ROC curve (AUC) was calculated for the combined metabolite panel adjusted for age, body mass index, smoking status, high-density lipoproteins and triglycerides.

4.5 Replication analysis in two independent cohorts

Replication analysis of the most significant ten alcohol-related metabolites in males and five metabolites in females found in KORA F4 discovery sample was performed in two independent KORA F3 and TwinsUK cohorts. In males, three out of ten metabolites (i.e. PC aa C32:1, PC aa C36:1, SM (OH) C16:1) could be replicated in KORA F3 (**Table 6**). In females, two out of five metabolites could be replicated (**Table 7**). One metabolite in KORA F3 (i.e. PC ae C30:2) and one metabolite (i.e. PC aa C34:1) in TwinsUK. In the TwinsUK population, only females were available for

replication analysis. 629 TwinsUK participants met the inclusion criteria and were eligible for the replication analysis, however only for 277 participants HDL and triglyceride data was available for the same time point. In TwinsUK the replication analysis was performed using 277 and 629 study participants. In the first replication analysis on 277 participants, logistic regression adjusted for age, BMI and smoking, HDL and triglyceride resulted in no significant p-values. Once the sample size was increased to 629 and the logistic regression model adjusted for age, BMI and smoking, was used, the metabolite PC aa C34:1 could be replicated. Additionally, data was pooled from the KORA F4 discovery and KORA F3 replication samples and a meta-analysis with a fixed effect model conducted in order to investigate the combined effect of alcohol on metabolite concentrations. In the meta-analysis the replication succeeded for all ten metabolites in men and five metabolites in women. This indicates that due to the small sample size in TwinsUK and KORA F3 cohorts the previous replication could not be achieved for all metabolites. Nevertheless, the trends of metabolite concentrations (as stated by the comparison of means of metabolite concentrations between MHD and LD in **Table 6** and **Table 7**) for all ten and five metabolites are consistent with the trends in the discovery across all studies. For example, the metabolite lysoPC a C18:1 was not replicated in KORA F3 and TwinsUK, still the mean metabolite concentration is higher in MHD compared to LD throughout the KORA F4, KORA F3 and TwinsUK studies.

Metabolite	Discovery KORA F4 (n=1144)			Replication KORA F3 (n=154)			Meta-analysis Discovery + Replication fixed effects (n=1298)		
	LD ^e mean±SD (μM) ^a	MHD ^f mean±SD (μM) ^a	P-value ^{b,c}	LD mean±SD (μM) ^a	MHD mean±SD (μM) ^a	P-value ^{b,d}	LD mean±SD (μM) ^a	MHD mean±SD (μM) ^a	P-value ^{b,d}
PC aa C32:1	19.3±9.2	31.0±16.9	1.1E-18	20.7±9.7	35.0±16.5	7.3E-05	19.4±9.2	31.6±16.8	8.6E-23
PC aa C36:5	28.7±12.9	36.6±17.6	3.9E-07	31.8±19.8	44.0±278	4.9E-02	29.0±13.9	37.7±19.5	5.6E-08
PC aa C40:4	4.0±1.2	4.9±1.6	1.2E-07	3.7±1.0	4.7±1.5	1.6E-02	4.0±1.2	4.8±1.6	2.8E-09
PC aa C36:1	51.4±11.9	59.7±16.4	6.6E-06	53.0±11.2	66.6±15.9	9.9E-04	51.5±11.8	60.7±16.5	9.9E-09
lysoPC a C17:0	1.7±0.5	1.5±0.5	5.4E-10	1.6±0.4	1.3±0.4	3.4E-02	1.7±0.5	1.5±0.5	4.1E-11
lysoPC a C18:1	20.0±5.5	22.7±7.4	1.6E-05	18.5±4.9	20.9±5.0	1.3E-01	19.9±5.4	22.5±7.1	3.9E-06
PC ae C40:6	4.9±1.2	4.4±1.1	4.3E-10	5.4± 1.3	5.0±1.1	3.5E-02	4.9±1.2	4.5±1.1	1.5E-10
SM(OH) C16:1	3.1±0.7	2.7±0.7	6.8E-12	2.7±0.7	2.4±0.6	2.0E-03	3.0±0.7	2.6±0.7	1.2E-13
SM(OH) C22:1	12.6±2.5	11.4±2.9	3.4E-09	10.0±2.5	9.3±2.1	1.2E-02	12.3±2.6	11.1±2.9	4.4E-10
SM(OH) C22:2	10.2±2.1	9.1±2.3	4.0E-13	8.1±2.1	7.7±1.7	2.3E-02	9.9±2.2	8.9±2.3	1.3E-13

Table 6. Result of logistic regression analysis of alcohol-specific key metabolites in males. Abbreviations: SD= standard deviation; LD= light drinkers; MHD= moderate-to-heavy drinkers' ^a mean and standard deviation of the metabolite concentration from serum, ^b logistic regression analysis adjusted for age, BMI, smoking, HDL, triglycerides, ^c significance level < 0.00038 (Bonferroni corrected), ^d significance level < 0.005 males (Bonferroni corrected), ^e alcohol consumption < 40g/day males, < 20g/day females, ^f alcohol consumption ≥ 40g/day males, ≥ 20 g/day females' Significant P-values are represented in **bold**.

Metabolite	Discovery KORA F4 (n=946)			Replication KORA F3 (n=107)			Meta-analysis Discovery + Replication fixed effect (n=1053) ^g			Replication TwinsUK (n=277)		
	LD ^e mean±SD (μ M) ^a	MHD ^f mean±SD (μ M) ^a	P-value ^{b,c}	LD mean±SD (μ M) ^a	MHD mean±SD (μ M) ^a	P- value ^{b,d}	LD mean±SD (μ M) ^a	MHD mean±SD (μ M) ^a	P- value ^{b,d}	LD mean±SD (μ M) ^a	MHD mean±SD (μ M) ^a	P- value ^{b,d}
PC aa C34:1	241.1±52.9	259.4±56.3	1.0E-04	274.8±72.8	308.5±73.6	5.8E-02	244.6±56.2	263.7±59.6	6.6E-05	311.5±98.8	350.4±121.0	9.4E-03 ^h
lysoPC a C16:1	2.8±0.9	3.1±1.0	4.7E-05	2.5±1.0	3.04±0.9	5.2E-02	2.7±0.8	3.1±1.0	6.2E-06	4.21±1.5	4.24±1.7	7.3E-01
lysoPC a C17:0	1.8±0.5	1.6±0.5	2.8E-04	1.6±0.5	1.4±0.4	2.3E-02	1.7±0.5	1.6±0.4	7.7E-05	2.3±0.9	2.2±0.7	5.6E-01
PC ae C30:2	0.17±0.04	0.16±0.04	4.0E-05	0.17±0.03	0.15±0.04	2.5E-03	0.16±0.04	0.15±0.03	4.2E-06	0.2±0.1	0.2±0.1	7.8E-01
PC ae C40:6	5.4±1.3	5.0±1.2	2.4E-07	6.1±1.5	5.4±1.4	2.2E-02	5.4±1.3	5.0±1.2	1.7E-08	7.5±2.7	7.4±2.7	9.9E-01

Table 7. Result of logistic regression analysis of alcohol-specific key metabolites in females. Abbreviations: SD, standard deviation; LD=light drinkers; MHD=moderate-to-heavy drinkers, ^amean and standard deviation of the metabolite concentration from serum, ^blogistic regression analysis adjusted for age, BMI, smoking, HDL, triglycerides, ^csignificance level < 0.00038 (Bonferroni corrected), ^dsignificance level < 0.01 females (Bonferroni corrected), ^ealcohol consumption < 40g/day males, < 20g/day females, ^falcohol consumption ≥ 40g/day males, ≥ 20 g/day females, ^gmeta-analysis consist of KORA F4 discovery, KORA F3 and TwinsUK replication data sets, ^hlogistic regression analysis adjusted for age, BMI, smoking with n=629 study participants. Significant P-values are represented in **bold**.

4.6 Characterization of the relationship between alcohol consumption and key metabolite concentrations

The previous global analysis using heat maps and cluster analysis revealed consistent clusters and trends in metabolite concentrations with increasing alcohol consumption for the set of 40/18 significant metabolites in males/females (**Figure 9**). A more detailed analysis of the trends and concentrations changes for the 10/5 key metabolites in males/females was conducted, especially to display detailed (“zoom in perspective”) fluctuations in metabolite concentrations with alcohol consumption. An analysis using smoother plots was used to provide deeper insights into those fluctuations. Smoother plots were drawn for the 10/5 key metabolites identified. As already displayed in the heat map representation, some metabolites showed increasing and some decreasing trend with increasing alcohol consumption. However, at low alcohol consumption (i.e. 0-40g/day/ 0-20g/day alcohol consumption in males/females) smoother plots displayed several major fluctuations of metabolite concentrations whereas in the moderate-to-high alcohol consumption range (i.e. threshold of >40g/day/>20g/day for males/females) (**Figure11**) (**Figure12**) smoother plots displayed a more linear association of metabolite concentration with alcohol consumption. Overall the results show a non-linear relationship between alcohol consumption and metabolite concentrations in (threshold effects)

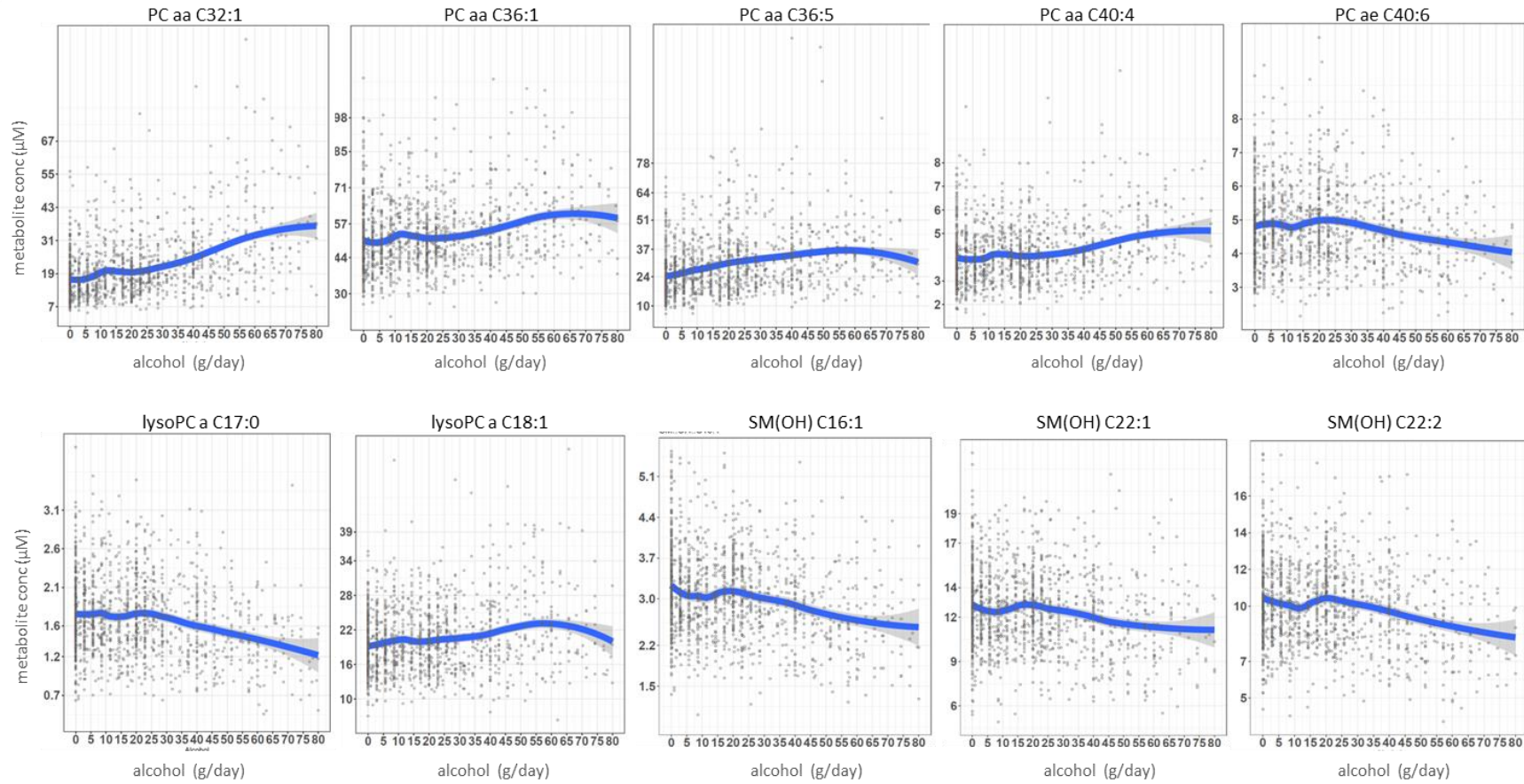


Figure 11. Non-linear relationship between alcohol consumption and metabolite concentrations in males. Smoother plots of 10 key metabolites in males (KORA F4). The plots show concentration trends of the 10 identified markers across age in males with age on the x-axis and concentration on the y-axis, respectively.

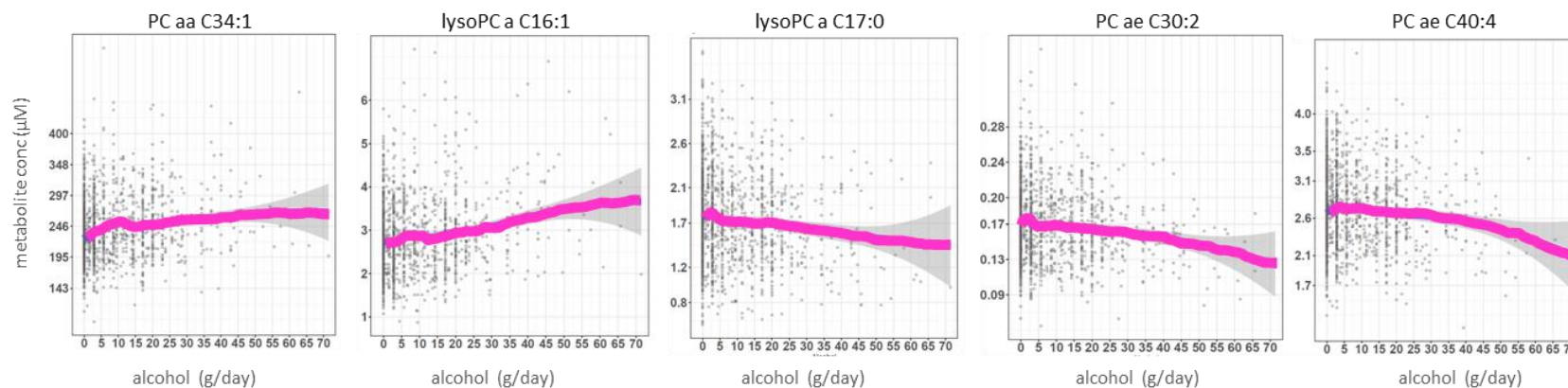


Figure 12. Non-linear relationship between alcohol consumption and metabolite concentrations in females. Smoother plots of 5 key metabolites markers in females. The plots show concentration trends of the 5 identified markers across age in KORA F4 males with age on the x-axis and concentration on they-axis, respectively.

4.7 Correlation of key metabolites and laboratory parameters (HDL and triglycerides)

HDL and triglycerides were previously used as covariates for adjusting in the logistic regression model. In order to investigate more closely the dependencies regarding the increasing/decreasing direction of HDL, triglycerides with increasing/decreasing metabolite concentrations and derive potential biochemical mechanisms of alcohol-induced metabolic profiles a correlation analysis between HDL, triglyceride and the 10/5 key alcohol-related metabolites in males/females was performed. In females, four out of the five key metabolites (i.e. PC aa C 34:1, lysoPC a 16:1, PC ae C 32:0, PC ae C40:4) were positively correlated with HDL and no correlation with HDL was observed for lysoPC a 17:0. Two out of five metabolites (i.e. PC aa C34:1, lysoPC a 16:1) were positively correlated with triglycerides, and two metabolites (i.e. lysoPC a C17:0, PC ae C40:4) were negatively correlated with triglycerides. No correlation with triglycerides was observed for PC ae C30:2 (**Figure13a**). In males, eight out of ten metabolites (i.e. PC aa C32:1, PC aa C36:1, PC aa C36:5, lysoPC a 17:0, lysoPC a C18:1, PC ae C40:6, SM(OH) C16:1, SM(OH) C22:1, SM(OH) C22:2) were positively correlated with HDL. No correlation was observed for PC aa C40:4. Four out of ten metabolites (i.e. PC aa C32:1, PC aa C36:1, PC aa C36:5, PC aa C40:4) were positively correlated with triglycerides, and six metabolites (i.e. lysoPC a C17:0, lysoPC a C18:1, PC ae C40:6, SM(OH) C16:1, SM(OH) C22:1, SM(OH) C22:2) were negatively correlated with triglycerides (**Figure13b**).

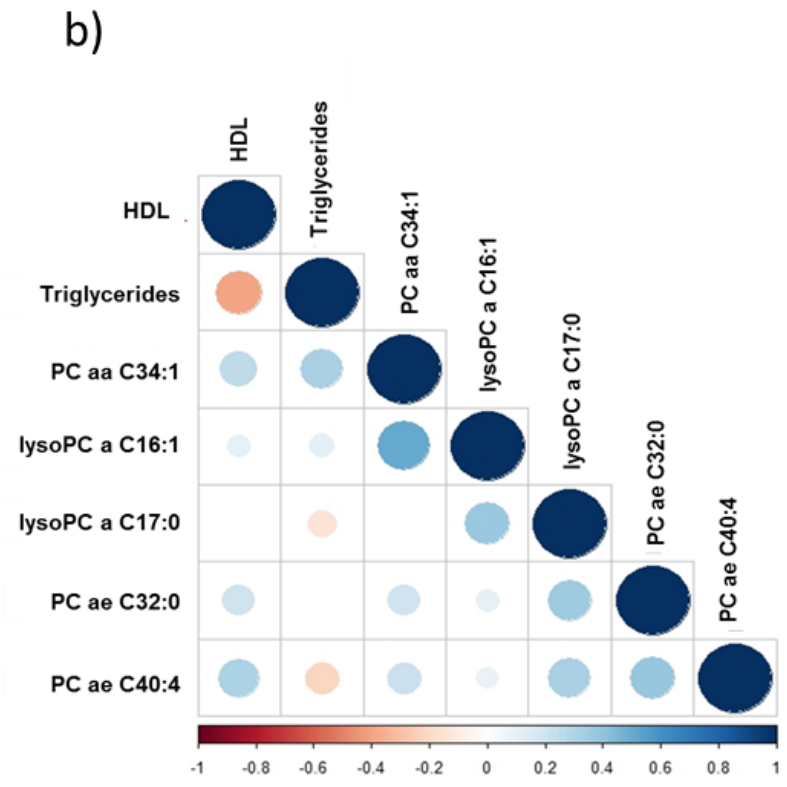
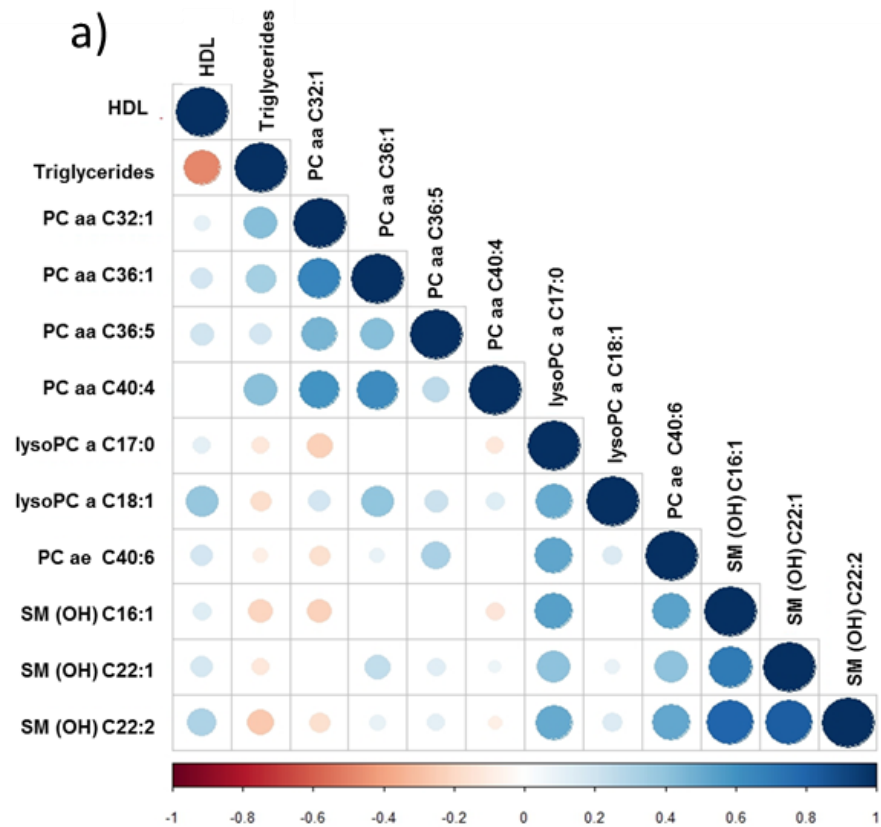


Figure 13. Correlation between 10/5 key metabolites in males/females and laboratory parameters HDL and triglycerides

Plot shows output of correlation analysis **(a)** between HDL, triglycerides and each the 10 most significant metabolites in males and **(b)** HDL, triglycerides and 5 most significant metabolites in females previously identified by step-wise selection and random forest analysis. Correlation was performed according to Spearman. In the plots, each cell represents a correlation coefficient (Spearman's rho). The correlation coefficients are represented by colored bubbles. Red color displays negative correlation, blue color displays positive correlation. No entry means no significant correlation observed. Rho = 1 corresponds to perfect positive correlation, rho=-1 corresponds to perfect negative correlation, rho=0 corresponds to no correlation

4.8 Alcohol consumption effects on metabolic pathways

In order to investigate which metabolic pathways are affected by alcohol consumption, metabolic pathway mapping of the five key alcohol-specific metabolites in females (i.e. PCaaC34:1, lysoPC a C16:1, lysoPC a C17:0, PC ae C30:2, PC ae C30:2, PC ae C40:6) and ten key metabolites in males (i.e. PC aa C32:1, PC aa C36:5, PC aa C40:4, PC aa C36:1, lysoPC a C17:0, lysoPC a C18:1, PC ae C40:6, SM(OH) C16:1, SM(OH) C22:1, SM(OH) C22:2) was performed against available human KEGG pathways. Over-representation analysis was performed to test if a group of compounds is represented more than expected by chance within the user uploaded compound list. In the context of pathway analysis, this means that we are testing, if compounds involved in a particular pathway are enriched compared by random hits. The pathway analysis showed that moderate-to-heavy alcohol consumption has the biggest impact on the Glycerophospholipid metabolism, Ether lipid metabolism and Sphingolipid metabolism. The results from the pathway analysis are presented graphically (**Figure 14**) as well as in a detailed table (**Table 8**)

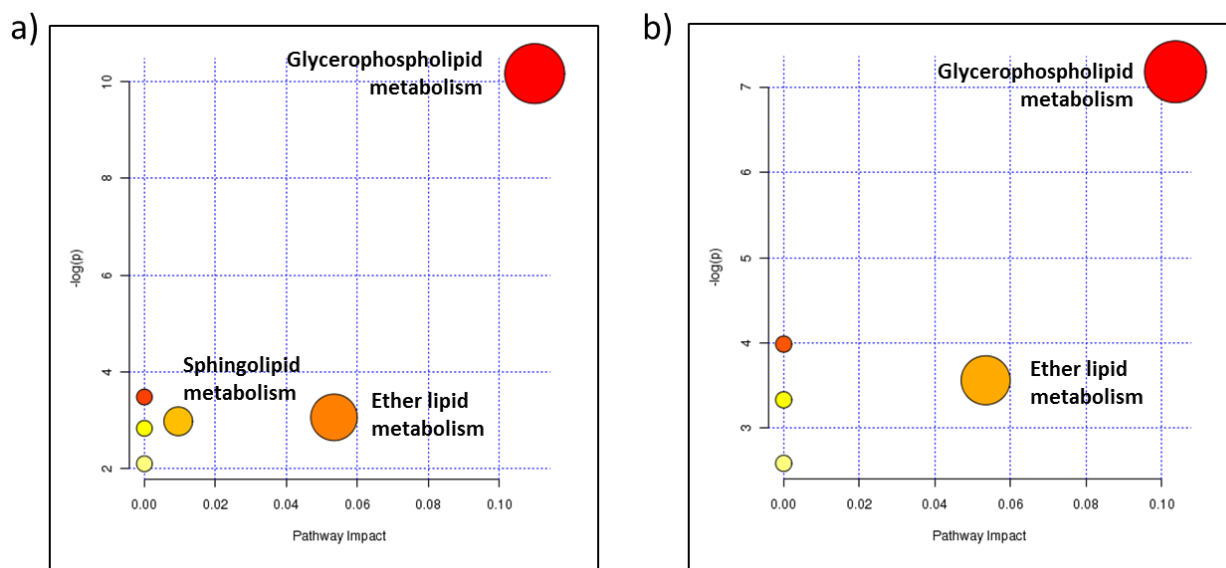


Figure 14. Summary of pathway analysis: Moderate-to-heavy alcohol consumption has the biggest impact on the Glycerophospholipid metabolism, Ether lipid metabolism and Sphingolipid metabolism. Schematic is adapted from the automatically generated Metaboanalyst 2.0 results report. The pathway analysis consists of enrichment and impact of alcohol-related metabolites in Kyoto Encyclopedia of Genes and Genomes pathways (KEGG) for (a) males and (b) females. The enrichment scores $-\log(p)$ are shown on y-axis, which was calculated as the negative logarithm of the P-value from an enrichment test-axis represents the Pathway Impact, scored between 0 and 1, which indicates the pathway importance of the metabolites.

a)

	Total	Expected	Hits	Raw p	$-\log(p)$	Holm adjust	FDR	Impact
Glycerophospholipid metabolism	39	0.08	3	3.85E-05	4.41E+00	3.08E-03	3.08E-03	0.11
Linoleic acid metabolism	15	0.03	1	3.08E-02	1.51E+00	1.00E+00	9.42E-01	0.00
Ether lipid metabolism	23	0.05	1	4.69E-02	1.33E+00	1.00E+00	9.42E-01	0.05
Sphingolipid metabolism	25	0.05	1	5.09E-02	1.29E+00	1.00E+00	9.42E-01	0.01
alpha-Linolenic acid metabolism	29	0.06	1	5.89E-02	1.23E+00	1.00E+00	9.42E-01	0.00
Arachidonic acid metabolism	62	0.13	1	1.22E-01	9.12E-01	1.00E+00	1.00E+00	0.00

b)

	Total	Expected	Hits	Raw p	$-\log(p)$	Holm adjust	FDR	Impact
Glycerophospholipid metabolism	39	0.05	2	7.60E-04	3.12E+00	6.08E-02	6.08E-02	0.10
Linoleic acid metabolism	15	0.02	1	1.86E-02	1.73E+00	1.00E+00	7.15E-01	0.00
Ether lipid metabolism	23	0.03	1	2.84E-02	1.55E+00	1.00E+00	7.15E-01	0.05
alpha-Linolenic acid metabolism	29	0.04	1	3.57E-02	1.45E+00	1.00E+00	7.15E-01	0.00
Arachidonic acid metabolism	62	0.08	1	7.53E-02	1.12E+00	1.00E+00	1.00E+00	0.00

Table 8. Detailed results from Pathway analysis. Table is adapted from the automatically generated Metaboanalyst 2.0 results report. Shows the enrichment and impact scores of alcohol-related metabolites in KEGG pathways. The $-\log(p)$ was considered as the enrichment score. Impact, scored between 0 and 1, indicates the pathway topological importance of the metabolites. In particular, the parameter Total is the total number of compounds in the pathway; the parameter Hits is the actually number of metabolites with significant variations in the pathway; the Raw p is the original p-value calculated from the enrichment analysis; the false discovery rate FDR are calculated by the p values adjusted using Benjamin-Hochberg method.

4.9 Alcohol-related metabolites and associated disease states & phenotypes

A systematic PubMed literature review of targeted metabolomics publications identified that the alcohol-related metabolites are also associated with other phenotypes and disease states (**Table 9**). Ten out of 13 metabolites were also reported in other studies. Three metabolites were exclusively found to be associated with moderate-to-heavy alcohol consumption in the current alcohol study (i.e. lysoPC a C16:1, SM(OH) C16:1, SM(OH) 22:1). Four out of 13 alcohol-related metabolites also occur in smokers (Xu, Holzapfel et al. 2013) and are associated with Diabetes risk (i.e. PC aa C 32:1, PC aa C36:1, PC ae 40:6, PC aa C40:4) (Floegel, Stefan et al. 2013). PC aa C 35:6 is also associated with Huntington's disease (HD) (Mastrokoulas, Pool et al. 2016), lyso PC a C18:1 with obesity (Wahl, Yu et al. 2012), PC aa 40:4 and SM(OH) 22:2 with Ovarian endometriosis (Vouk, Hevir et al. 2012) and PC ae C30:2 with Prostate cancer (Schmidt, Fensom et al. 2017).

Metabolite	Gender	Alcohol consumption i.e. moderate-to-heavy	Smoking i.e. current smoker	Red meat consumption	Obesity i.e. child obesity	Diabetes	Ovarian endometriosis	Huntington's disease	Prostate cancer	References
PC aa C36:5		↑*						↓		(Jaremek, Yu et al. 2013) (Mastrokolias, Pool et al. 2016)
lysoPC a C16:1		↑**								(Jaremek, Yu et al. 2013)
SM (OH) C16:1		↓*								(Jaremek, Yu et al. 2013)
SM (OH) C22:1		↓*								(Jaremek, Yu et al. 2013)
PC aa C34:1		↑**	↑*							(Jaremek, Yu et al. 2013) (Xu, Holzapfel et al. 2013)
PC aa C32:1		↑*	↑**			↑				(Jaremek, Yu et al. 2013) (Xu, Holzapfel et al. 2013) (Floegel, Stefan et al. 2013)
PC aa C36:1		↑*	↑** ↑**			↑				(Jaremek, Yu et al. 2013) (Xu, Holzapfel et al. 2013) (Floegel, Stefan et al. 2013)
PC ae C40:6		↓* ↓**	↓*			↓				(Jaremek, Yu et al. 2013) (Xu, Holzapfel et al. 2013) (Floegel, Stefan et al. 2013)
PC aa C40:4		↑*	↑*				↓**			(Jaremek, Yu et al. 2013) (Xu, Holzapfel et al. 2013) (Vouk, Hevir et al. 2012)
lysoPC a C17:0		↓* ↓**		↓		↓				(Jaremek, Yu et al. 2013) (Wittenbecher, Muhlenbruch et al. 2015)
lysoPC a C18:0	↑*	↑*			↓					(Jaremek, Yu et al. 2013) (Mittelstrass, Ried et al. 2011) (Wahl, Yu et al. 2012)
PC ae C30:2	↑**	↓**							↑	(Jaremek, Yu et al. 2013) (Mittelstrass, Ried et al. 2011) (Schmidt, Fensom et al. 2017)
SM (OH) C22:2	↑**	↓*	↓**				↑**			(Jaremek, Yu et al. 2013) (Mittelstrass, Ried et al. 2011) (Xu, Holzapfel et al. 2013) (Vouk, Hevir et al. 2012)

Table 9. Alcohol-related key metabolites are also associated with other phenotypes and diseases states. Table columns describe the phenotype under investigation as e.g. gender, smoking and alcohol consumption (i.e. current metabolomics study); (**) indicates separate analyses performed only on the female population; (*) analyses only on male population, (↓) indicates a decrease of metabolite concentration; (↑) indicates an increase in metabolite concentration reported in the study.

5. Discussion

In the following chapter the study design, data quality, metabolic panel, underlying potential metabolic pathways and association with medical conditions as well as the relevance of biomarkers and the given mass spectrometry setup for alcohol clinical diagnostics are discussed.

5.1 Metabolomics study

5.1.1 Study design

The underlying work is based on a population-based case-control study design with three independent and not related populations from Germany and UK. Advantage of the study design is the large size of discovery sample enabling a high statistical power. Limitations occur in the replication of the results due to small replication data sets. Still, the results can be considered as successfully replicated.

5.1.2 Data quality

Alcohol consumption was measured by a questionnaire that investigates the intake throughout the proband's life. Alcohol intake was characterized in g/per day and was calculated by averaging the intake of beer, wine and liquor, reported elsewhere (Ruckert, Heier et al. 2011). An error in estimating alcohol consumption is unavoidable using such a method of investigation due to the variability in averaging alcohol intake for long consumption periods. In our study, this potential bias can be considered reasonably negligible, since, characterized by a steady daily use allowed us to obtain realistic data on alcohol intake.

5.1.3 Study populations (Alcohol consumption groups)

For Drinkers no information was available on the duration of alcohol consumption of

drinkers (i.e. long life-drinkers or recent drinkers). For Abstainers data was not available to the level of granularity to study potential metabolic differences between different abstainer groups. An interesting aspect that could not be covered in this study but relevant for further investigations would be the comparison of metabolic profiles between recent abstainers and long-term abstainers and life-time abstainers.

5.1.4 Methodology to identify significant metabolites

In this study a hypothesis-driven approach was used to investigate significant metabolite differences between predefined alcohol consumption groups. A comparison of absolute metabolite concentrations between Abstainers, Low-Drinkers (LD) and Moderate-to-Heavy Drinkers (MHD) was not performed as a first step of this candidate biomarker discovery due additional factors for alcohol that need to be taken into consideration that's why a test, as for example, Kruskal-Wallis to identify significant metabolites based only on the information of metabolite concentrations as the single point of difference not taking into consideration other factors as lifestyle was not chosen. An analysis using non-hypothesis- driven methods on the set of 131 metabolomics data using e.g. Volcano plots or self-organizing maps (SOM), Partial-Least Square Methods (PLS-DA) did not lead to results (data not displayed) i.e. no patterns or clusters could be derived. Data was not available to investigate correlations regarding other alcohol-related and relevant confounders as inflammation markers as CRP, traditional alcohol-related biomarkers/liver biomarkers e.g. GGT or genetic data to perform gene expression, or SNP data to study genotype effects on metabolite concentrations.

5.1.5 Metabolic Pathway analysis

Metabolite names and descriptions are often not harmonized across different

databases. An example is KEGG: the granularity of KEGG pathways is suboptimal. Different metabolites have the same KEGG ID e.g. SM (OH) C16:1; SM (OH) C22:1; SM (OH) C22:2. KEGG does not necessarily makes the distinction between different chain length of e.g. Sphingomyelins. Different metabolite subtypes are interpreted as same. Additional and laborious analyses needed to be performed using publication desktop research and other platforms as e.g. HMDB. Urgently a standardization is needed of metabolomics database platforms.

5.2 Alcohol-related metabolites

5.2.1 Metabolomic candidate biomarker panel and its potential alcohol-related underlying biochemical mechanisms

In the current study a targeted metabolomics approach was used that identified, as well as partly replicated alcohol-related metabolites in German and UK human studies. The results suggest that alcohol affects mostly the sphingolipid, glycerophospholipid and ether lipid metabolism. A schematic overview of the observed alcohol-specific metabolic differences and the potential underlying mechanisms is depicted in **(Figure 15)** and are discussed below.

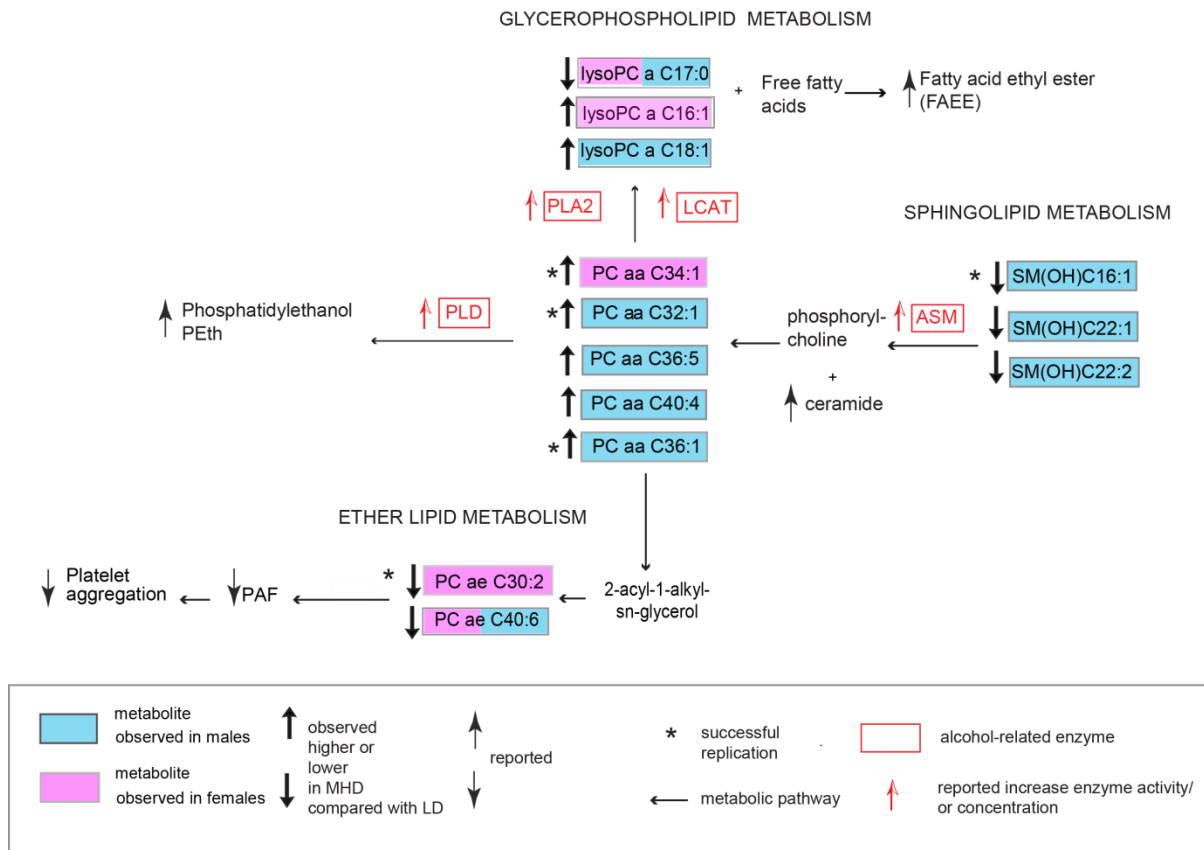


Figure 15. Potential alcohol-induced biochemical mechanisms. Schematic overview of metabolite concentration differences in MHD compared to LD in males and females. Ten/five metabolites that best discriminate MHD from LD in males/females are shown. Pink and blue boxes represent male- and female-specific alcohol-related metabolites identified in this study. Combined pink-blue boxes represent metabolites identified both in males and females. Bold black arrows represent observed higher or lower of metabolite concentration in MHD compared with LD in the discovery. Replicated metabolites are marked by a star (*). Thin black arrows represent the higher or lower of alcohol-related analytes in MHD reported in earlier publications. Red boxes represent alcohol-related enzymes and red arrows represent the effect on the respective enzyme activity or concentration reported in previous publications in MHD. males/females are shown.

Spingomyelins

The underlying mechanism for lower sphingomyelin concentrations (SM(OH)C16:1, SM(OH)C22:1, SM(OH)C22:1) in MHD compared with LD could be attributed to (ASM) acid sphingomyelinase activity. ASM catalyzes the hydrolysis of sphingomyelins by cleaving the phosphodiester bond of sphingomyelins generating ceramide and phosphorylcholine (Liu, Wang et al. 2000, Jenkins, Canals et al. 2009) which is again reassembled to phosphatidylcholine (Li, Hailemariam et al. 2007).

Enzymatic dysfunction of ASM results in Niemann-Pick disease A (NPD-A, OMIM 257200) and B (NPD-B, OMIM 607616) a lipid storage disease characterized by accumulation of sphingomyelins within the endo-lysosomal compartment (Jenkins, Canals et al. 2009). Interestingly, this mechanism is reciprocal when alcohol is administered. Several studies investigating cellular response to alcohol in-vitro and in-vivo have provided evidence that alcohol stimulates the ASM activity leading to accumulation of ceramide and decrease of sphingomyelins (Deaciuc, Nikolova-Karakashian et al. 2000, Liu, Wang et al. 2000, Pascual, Valles et al. 2003, Saito, Saito et al. 2005, Viktorov and Yurkiv 2008). A recent *in-vivo* study on patients with alcohol dependence reported alcohol-induced release of phosphorylcholine from sphingomyelins in peripheral blood cells confirming alcohol-induced activation of ASM (Reichel, Greiner et al. 2010).

Diacylphosphatidylcholines

There is a direct correlation between diacylphosphatidylcholine (PC) concentrations and phosphatidylethanol (PEth). PEth is a clinical biomarker of the past 1-2 weeks of moderate-to-heavy alcohol consumption (Stewart, Reuben et al. 2009). PEth is a unique phospholipid that is synthesized only in the presence of ethanol and is directly formed from PCs by the enzyme phospholipase D (PLD) (Gustavsson and Alling 1987, Kobayashi and Kanfer 1987, Mueller, Fleming et al. 1988) that catalyzes the exchange of ethanol for choline in PCs (Mueller, Fleming et al. 1988). Different PEth molecular species have a common phosphoethanol head group onto which two fatty acid moieties derived from PCs are attached. A study by Helander et al. (Helander and Zheng 2009) has shown that PEth-16:0/18:1 (34:1) was the most predominant molecular species accounting for 37% of all PEth species. A recent study by Nalesso et al. (Nalesso, Viel et al. 2011) compared the occurrence of

different PEth species between heavy drinkers and social drinkers (defined as daily alcohol intake 60-300g/day and 0-20g/day respectively). Interestingly, PEth 16:0/18:1 (34:1), PEth 18:0/18:1 (36:1) and PEth 16:0/16:1 (C32:1) were most abundant in heavy drinkers. This may be consistent with our findings in which PC aa C34:1 in female, PC aa C36:1 and PC aa 32:1 in male had higher concentration MHD compared to LD. We hypothesize that concentrations of specific PC species can be used as surrogate biomarkers for PEth to distinguish MHD from LD. However, PEth measurements are out of scope of this study. Dedicated and parallel measurements of PC aa C34:1 and PEth (34:1) would be required in order to investigate whether PC aa C 34:1 can be a substitute PEth (34:1).

Lysophosphatidylcholines

Lysophosphatidylcholines (lysoPCs) are derived from PCs (Croset, Brossard et al. 2000) and have been reported to have cytotoxic effects (Weltzien 1979). They accumulate in alcohol-related conditions as in atherosclerosis (Matsumoto, Kobayashi et al. 2007) or ischaemia (Wang, Wang et al. 2001). LysoPCs originate from several metabolic pathways, as part of the production is attributed to the transesterification of PCs and free cholesterol catalysed by the enzyme lecithin-cholesterol acyltransferase (LCAT), where LCAT hydrolyses the sn-2 acyl group and subsequently transfers and esterifies the fatty acid to free cholesterol (Holleboom, Kuivenhoven et al. 2010). A study by Goto et al. (Goto, Sasai et al. 2003), investigating clinical alcoholics, reported an increase of LCAT concentration in individuals with alcohol intake of >30g/day. Another metabolic pathway generating LysoPC species is attributed to the enzyme phospholipase A2 (PLA2), which catalyzes the hydrolysis of an ester bond at the sn-2 position of 1,2-sn-diacylglycerols yielding lysoPCs and free fatty acids (Caro and Cederbaum 2006)

which are esterified into fatty acid ethyl esters (FAEE) that have been reported as alcohol marker to distinguish social from heavy drinkers or alcohol-dependent individuals (Salem, Refaai et al. 2001, Wurst, Alexson et al. 2004). Fatty acids with uneven number of carbons (i.e. C15:0 and C17:0) are produced by bacterial flora of human intestine (Hopkins, Sharp et al. 2001). It is known that alcohol acts as a disinfectant which kills bacteria. Thus, a possible explanation for the lower concentrations of lysoPC17:0 in MHD could be that alcohol consumption leads to the disruption of the respective intestinal bacterial microflora in the gut which thus influences lysoPC a C17:0 levels in human blood. On the other hand, the fatty acid C17:0 is also found in the bacterial flora of ruminants (Wu and Palmquist 1991, Smedman, Gustafsson et al. 1999). A study by Wolk et al. (Wolk, Vessby et al. 1998) revealed that portions of the fatty acids C15:0 and C17:0 in adipose tissue reflected milk fat consumption in women. An earlier study (Ruf, Nagel et al. 2005) investigating associations of reported alcohol intake with dietary habits in probands from the EPIC cohort found that alcohol consumers had a lower intake of dairy products than abstainers. This is consistent with another French cohort of the EPIC study (Kesse, Clavel-Chapelon et al. 2001) which found that high alcohol intake was associated with lower consumption of dairy products in both genders compared to moderate alcohol consumption. Thus, another plausible explanation to the lower concentrations of lysoPC17:0 in MHD in our study could be based on lower intake of dairy products. Based on the above findings and explanations, lysoPC a C17:0 might also be a dietary biomarker associated with distinguished dietary behavior of MHD compared to LD rather than a biomarker for alcohol-induced toxic or inflammatory mechanisms.

Ether lipids

Ether lipids (e.g. PC ae C30:2 and PC ae C40:6) play a role as precursor of platelet-activating factor (PAF)(Snyder 1995, Nagan and Zoeller 2001). PAF is an important mediator in hemostasis and it plays an important role in platelet aggregation (i.e. thrombotic effects). A number of studies indicate that ethanol directly affects hemostasis via a number of mechanisms, including platelet aggregation and activation(Mikhailidis, Jenkins et al. 1986, Hillbom and Neiman 1988, Dimmitt, Rakic et al. 1998, Salem and Laposata 2005). This mechanism is still not fully understood, however based on our results, it can be hypothesized that reduced PAF levels in response to moderate-to-heavy alcohol consumption might form a bottleneck in the process of platelet activation leading to poor platelet aggregation and to alcohol-related hemorrhagic events. This is supported by studies from the United States and Sweden showing that the baseline incidence of acute upper gastrointestinal bleeding increased by 3-fold as alcohol consumption increased from 1 drink or fewer to more than 20 drinks per week(Kaufman, Kelly et al. 1999).

5.2.2 Alcohol-related metabolites and their association with medical conditions

Diabetes

In the underlying study metabolites PC aa C32:1, PC aa C36:1, PC ae 40:6, and lysoPC a C17:0 were found elevated in moderate to-heavy drinkers. A prospective study by Floegel et al. (Floegel, Stefan et al. 2013) investigating the association between PC aa C32:1, PC aa C36:1, PC ae 40:6 and risk for Type 2 Diabetes found that those metabolites significantly improved T2D prediction compared with established risk factors and they were further linked to insulin sensitivity and secretion. According to Floegel et al. the study results indicate that PC aa 32:1, PC

aa 36:1, PC ae 40:6, are associated with early onset with higher risk of T2D. Another study by Wittenbecher et al. (Wittenbecher, Muhlenbruch et al. 2015) investigated blood metabolites that possibly relate red meat consumption to the occurrence of Type 2 Diabetes. The study reported decreased concentrations for lysoPC a C17:0 and identified it as associated with red meat consumption and diabetes risk. The association with red meat consumption would also confirm the hypothesis raised in this study (see chapter 5.2.1) that lysoPC a C17:0 could be a candidate biomarker of dietary behavior of moderate-to-heavy drinkers compared to low-drinkers rather than a biomarker for alcohol-induced toxic or inflammatory mechanisms.

Ovarian endometriosis

In the underlying alcohol study PC aa 40:4 was identified as elevated in moderate-to-heavy male drinkers and SM (OH) C22:2 was identified as decreased in moderate-to-heavy drinkers. Interestingly two studies by Vouk et al. (Vouk, Hevir et al. 2012, Vouk, Ribic-Pucelj et al. 2016) report the opposite effects for the ovarian endometriosis condition in females. The studies support the importance of phosphatidylcholine and sphingomyelin metabolites in the pathophysiology of endometriosis, however it is not clear how alcohol contributes into the condition of endometriosis as alcohol consumption was not part of the investigation.

Prostate cancer

In the underlying study PC ae C30:2 was decreased in female moderate-to-heavy drinkers. An opposite effect was reported in males for Prostate cancer (Schmidt, Fensom et al. 2017). Schmidt et al. investigated the prospective association between plasma metabolite concentrations and risk of prostate cancer overall, and by time to diagnosis and tumor characteristics, and risk of death from prostate cancer. Among several other metabolites PC ae C30:2 was associated with prostate

cancer and was inversely related to advanced stage of disease. As Schmidt et al. adjusted for alcohol consumption it can be hypothesized that PC ae C30:2 is also a real and independent proxy of prostate cancer.

Huntington's disease

In the underlying study PC aa C36:5 was identified as elevated in moderate-to-heavy male drinkers. Another study by Mastrokolias et al. (Mastrokolias, Pool et al. 2016) investigated metabolic changes in patients with Huntington's disease (HD) to identify disease related changes. Among several other metabolites in the study PC aa C36:5 exhibited significant association with disease severity supporting the notion that Phosphatidylcholine metabolism is also deregulated in HD.

5.3 Clinical utility of metabolic candidate biomarkers

5.3.1 Test scenarios in alcohol-related investigations

Depending on the diagnostic goal there are currently different test approaches. Alcohol intake a qualitative test that would categorize individuals into non-drinking, not abusive, and heavy drinking groups. The test would allow the identification of subjects with ongoing heavy drinking behaviors and could be applied in a hospital setting as a fast, multiparametric test for screening purposes (Freeman and Vrana (2010)), or as a differential diagnosis for non-alcohol induced liver damage or in monitoring of alcohol-withdrawal (National Institute for Health and Care Excellence 2011). Drinking patterns : a test of alcohol intake patterns that include periods of intoxication as even though total alcohol intake is the same over a week there is a difference between two drinks per day vs seven drinks per day in the same weekend (Freeman and Vrana (2010)). Alcohol-induced organ damage – a test that would identify early pre-symptomatic stages of disease development, would enable earlier and more effective treatment (Freeman and Vrana (2010)). By

detecting the onset and extent of organ damage, such diagnostics test could also provide an interventional “wake-up call” by the physician/therapist to heavy drinkers, and therefore enhance the biomarkers’ treatment utility through altering patients’ drinking behavior (Freeman and Vrana (2010)).

The underlying metabolomics study identified potential biomarkers that differentiate between low-drinkers and moderate-to-heavy-drinkers (**i.e. Biomarker of alcohol intake**) and to the current knowledge and outcomes of this study, the candidate biomarkers would be most qualified for the scenario to identify individuals with a chronic abusive drinking behavior. Based on the current research outcomes from other studies investigating the relationship between metabolites with medical conditions in (see chapters 4.9 and 5.2.2), those candidate biomarkers cannot be explicitly allocated to one particular pathophysiology leading to a particular organ damage, however the underlying results indicate that the alcohol-related biomarkers share a common biochemical pathway with Diabetes, Prostate Cancer, Huntington’s disease and Ovarian endometriosis.

5.3.2 Potential application areas

As stated in chapter 1.2.1, because patients may not disclose alcohol consumption or may underreport alcohol consumption, it is important to conduct alcohol biomarkers testing. Metabolomic alcohol biomarkers could be used in primary care (i.e. screening for alcohol abuse or potential relapse). Although a primary care physician currently does not routinely test for specific alcohol biomarkers in all patients, alcohol-specific metabolic changes during a routine laboratory investigation could alert a clinician regarding a potential chronic heavy drinking behavior. Kapoor et al. reported that when adding alcohol biomarker to a patient’s self-report of

alcohol consumption in the primary care setting could result in significant savings of healthcare costs (Kapoor, Kraemer et al. 2009). Another potential clinical setting would be the hospital (emergency) setting (i.e. screening for alcohol abuse or alcohol-related organ damage, alcohol-related disorder at admission). Gerke et al. reported that 20,9% of hospital inpatients were admitted due to alcohol-related disorders (Gerke, Hapke et al. 1997). Another study by Baune et al. investigated, which proportion of hospital admissions, which had been subject to inpatients with more than 11 emergency hospital admissions in the last 6 months in the region of Dortmund, were directly attributable to alcohol consumption. In total, 3% of all hospital inpatient admissions were caused alone by alcohol withdrawal syndrome and the average inpatient stay for these patients was 10 days (Baune, Mikolajczyk et al. 2005). An alcohol screening test for all patients at hospital admission could enable an early clinical decision making and shorten the follow-up diagnostics and improve therapy.

5.3.3 Mass spectrometry in the clinical lab

Mass spectrometry is superior to existing laboratory analysis methods (e.g. immunoassays) with respect to sensitivity, specificity and assay development time (Vogeser and Seger 2008). Mass spectrometry determines the mass-to-charge ratio (m/z) of particles allowing its accurate identification. It is an extremely sensitive technique and is very well suited for detecting small and large molecules at low concentrations (i.e. ng/ml). It can simultaneously measure hundreds of components present in complex biological media in a single assay. The biggest advantage is that there is no need for labeled reagents, thus, offering prospects of reduced time involved for assay development, simplified protocols, and lower cost of development.

Modern clinical laboratories use diverse techniques and instrumentation that vary in reliability and specificity. Since the introduction of mass spectrometry in clinical laboratories, it proved to be one of the most specific analytical techniques available for clinical diagnostics (Jannetto and Fitzgerald 2016). Mass spectrometry has proved as a successful technology in clinical Newborn Screening, Therapeutic Drug Monitoring (TDM) (e.g. antidepressants), Drugs of Abuse (DoA) (e.g. Methadone, Blood alcohol levels), Microbiology (i.e. Pathogen Identification) and Nucleic acid testing (i.e. SNP genotyping) applications and continues to expand in almost every area of laboratory medicine (Chace, DiPerna et al. 1999, Pusch, Wurmbach et al. 2002, Eichhorst, Etter et al. 2009, Wieser, Schneider et al. 2012, Desrosiers, Scheidweiler et al. 2015, McShane, Bunch et al. 2016) **(Figure 16), (Figure 17)**.

Method	Instrumentation	Application fields of laboratory medicine									
		Clinical chemistry & Immunoassay					Microbiology		Molecular/NAT		
		TDM	DoA	Newborn Screening	Endocrinology	Toxicology	(Protein) Path.ID	(Protein) Path.AST	Path.ID	Path.AST	human SNP genotyping
ESI-TOF	ESI-TOF								•	•	
	LC-ESI-TOF								•	•	
MALDI-TOF	MALDI-TOF						•				•
GC/MS	GC-EI-Quadrupole	•	•			•					
	GC-EI-Ion Trap	•	•			•					
	GC-CI-Quadrupole	•	•			•					
	GC-CI-Ion Trap	•	•			•					
LC/MS/MS	LC-ESI-Quadrupole/TOF	•	•		•	•					
	LC-ESI-Quadrupole/Quadrupole	•	•	•	•	•					
	LC-APCI-Quadrupole/Quadrupole	•	•	•	•	•					
	LC-ESI-Quadrupole/Ion Trap	•	•	•	•	•					
	LC-APCI-Quadrupole/Ion Trap	•	•	•	•	•					
	LC-ESI-Ion Trap/Ion Trap	•	•	•	•	•					

Figure 16. Mass spectrometry is used in almost all areas of laboratory medicine. Depicts the currently common applications in clinical laboratory. Schematic compiled and modified according to mass spectrometry industry market reports from SELECTBIO Mass Spectrometry for Clinical Diagnostics Market Report (2013), www.absciex.com, Laborjournal 1-2/2010 Geräte für Proteomics, www.directindustry.com

Method	Application	Analytes	Qualitative /Quantitative measurement	Sensitivity	Sample source	Example tests /Indication
GC/MS	DoA TDM	Volatile, low molecular compounds (< 1000 Da) as drug metabolites (e.g. cocaine : 303 Da)	qualitative quantitative	ng/mL - µg/mL (10E-9 – 10E-6)	urine whole blood serum plasma	e.g. Benzodiazepine e.g. Opiates : cocaine
LC-MS/MS	DoA TDM Detection of metabolic disorders Quantification of serum proteins	Mainly low molecular compounds (<1000 Da) as drug compounds, drug metabolites, amino acids, acylcarnitines, vitamins, hormones, steroids and oligosaccharides, peptides (>1000 Da)	qualitative quantitative	ng/mL (10E-9) > pg/mL (10E-12) (e.g. Testosterone)	urine whole blood serum plasma dried blood spots	e.g. Benzodiazepine, Tacrolimus, Sirolimus e.g. up to 40 disorders such as PKU, MCAD
MALDI - TOF	Microbial detection Genotyping to detect genetic disorders	High molecular compounds (> 1000 Da) as peptides, proteins, oligonucleotides	qualitative	N.A.	whole blood serum plasma urine swab smear	e.g. hundreds of bacterial strains and fungus types e.g. Cystic fibrosis testing (mutation in the CFTR gene)
ESI - TOF	Pathogen identification (bacteria, virus, fungus, yeast)	High molecular compounds (> 1000 Da), peptides, proteins, oligonucleotides (DNA, RNA)	qualitative	N.A.	whole blood serum plasma	e.g. E.coli, Chlamydia, Poliovirus

Figure 17. Mass spectrometry methods cover a wide spectrum of different analytes used in laboratory diagnostics.

Shows common methods specifications depending on the biomolecule class under investigation Schematic compiled and modified according to mass spectrometry industry market reports from SELECTBIO Mass Spectrometry for Clinical Diagnostics Market Report (2013), www.absciex.com, Laborjournal 1-2/2010 Geräte für Proteomics, www.directindustry.com.

5.3.3.1 From bench-to-bedside - key factors to consider for transition into clinical practice

Generally, stakeholders from the diagnostic industry, laboratory clinicians, primary care physicians, and health insurers will examine new methods/candidate biomarkers regarding their actual diagnostic test performance but also regarding the therapeutic consequences of the additional information they provide. When developing a new method, the overall test performance is essential for the routine clinical practice to deliver lab results on time, on budget in a quality. Despite the benefits of mass spec technology for the clinical laboratory some key additional aspects need to be considered e.g. : Automation in pre-analytics : at present mass spectrometry automation is rather limited in the clinical laboratory (in comparison to immunoassays or clinical chemistry analyzers) mainly due to the fact that the handling characteristics of mass spectrometry applications in the pre-analytics/sample preparation are still very far from the standards realized in routine clinical chemistry analyzers. Whereas common routine analyzer is fully automated, mass spectrometry applications still require significant hands-on-time due to the sample preparation **(Figure 18)**. Turn-around Time (TAT): of current mass spec applications for e.g. Newborn screening, DoA and TDM vary between minutes to hours **(Figure 19)**. More concisely, a potential mass spectrometry-based method for screening of abusive drinking behavior at hospital admission for all patients or at emergency setting would typically require a short TAT. Cost per analysis: cost per mass spec analysis varies from 0.1€ to 2€ and is lower that for Clinical chemistry/Immunoassay (CC/IA) methods which is (0.10€ - 7€). Sample throughput: for currently available mass spectrometry-based DoA /TDM tests is rather slow with 10 to 20 samples/hour compared to CC/IA methods with 1000 to 2000 samples/hour.

Maintenance/Consumables: High consumables usage, high degree of waste, open well-reagent stability, typical few weeks in CC/IA and extremely low usage of consumables, open-well reagent stability (~ 1 year) in mass spectrometry applications. Robustness: in mass spectrometry applications the daily measurement series up to 24h duration with short simple maintenance interventions possible whereas e.g. in CC/IA robustness is limited (Vogeser 2003, Vogeser and Seger 2008, Vogeser and Kirchhoff 2011, Vogeser and Seger 2012).

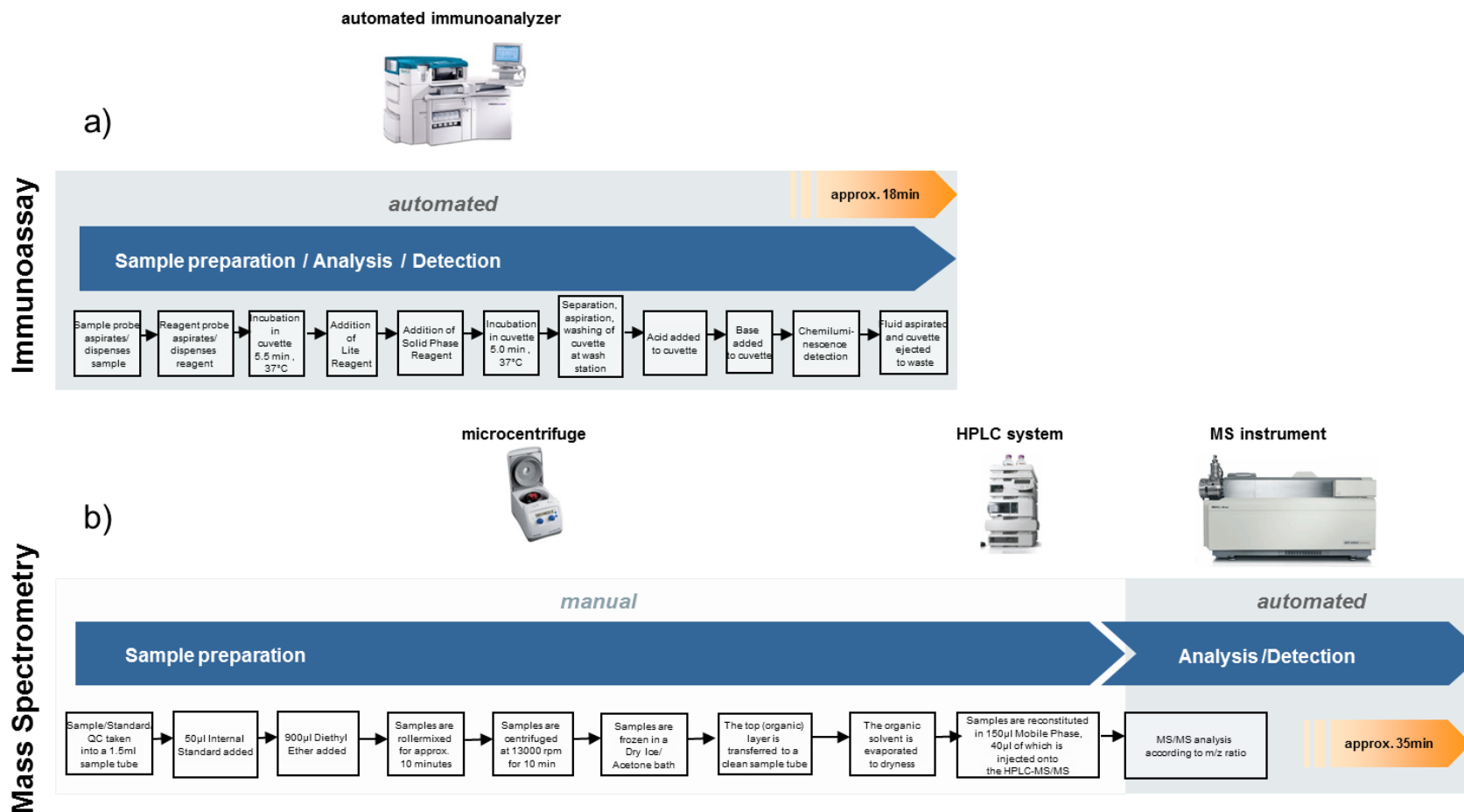


Figure 18. Mass spectrometry applications still require significant hands-on-time due to manual sample preparation. Displayed is an example laboratory workflow comparing an immunoassay with a mass spectrometry assay for Testosterone. **(a)** fully automated immunoassay is performed on a single instrument, starting with sample preparation, analysis and ending with automated waste disposal at a short turn-around-time i.e. “plug and play” **(b)** manual and time-consuming sample preparation dominates the mass spectrometry analysis. Beside the mass spec instrument additional lab instruments are required to perform an analysis. Schematic created and modified according to ADVIA Centaur CP/XP product manual (2010); Kushnir et al. Performance Characteristics of a Novel Tandem Mass Spectrometry Assay for Serum Testosterone (2006), Application note API 4000, ABSciex (2007)

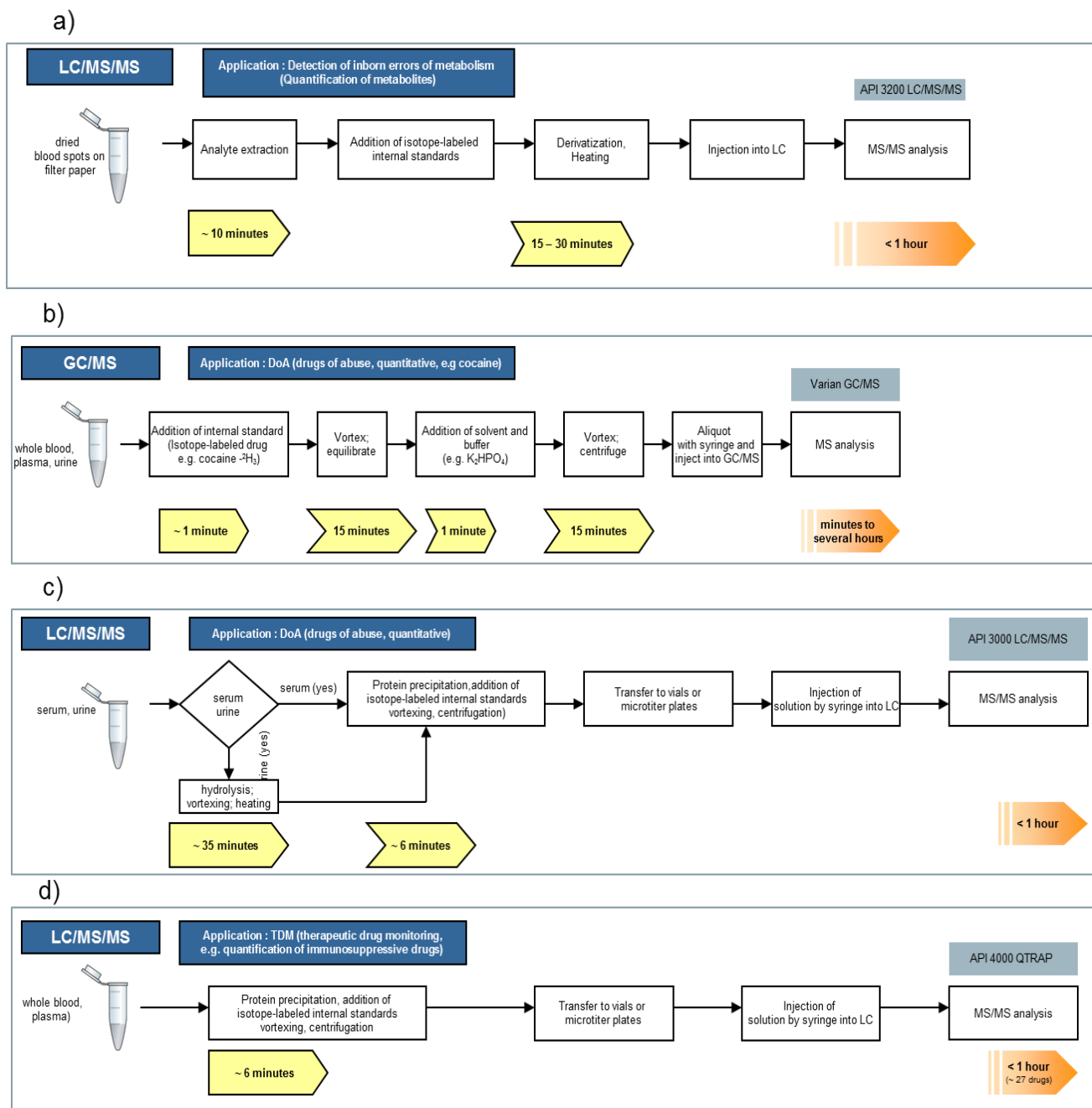


Figure 19. Current clinical laboratory processes with mass spectrometry still have a relatively long Turn-around-Time (TAT). Displayed are use cases of common laboratory workflows used in German clinical laboratories today. For each diagram the mass spec configuration is displayed (i.e. LC/MS/MS) and the Turn-around-time (TAT) for key analysis steps. Also, the overall TAT from sample-to-assay result is displayed. **(a)** Newborn Screening **(b)** and **(c)** Drugs of Abuse (DoA) and **(d)** Therapeutic Drug Monitoring (TDM). Schematics created and modified according to procedure descriptions at Labor Wisplinghoff, Köln (2012), Labor Limbach Karlsruhe (2012), SOPs für ABSciex and Varian Systems (2012), AB Sciex, SOPs for API 3200 LC/MS/MS System (2015)

5.3.4 What's next? - alcohol metabolomics study in the context of the biomarker development workflow

The conventional pipeline for biomarker development involves a discovery phase, followed by biomarker verification, validation, qualification and clinical application (Kitteringham, Jenkins et al. 2009). The focus of the underlying metabolomics study was to investigate the capability of targeted metabolomics to find metabolite differences of alcohol consumption groups, identify candidate biomarkers and identify potential application areas. This metabolomics study addresses the first phase of a biomarker development workflow (**Figure 20**) (**Figure 21**). In the underlying study a panel of candidate biomarkers could be identified and partly replicated. However, additional research is needed for the biomarker panel to be qualified for further steps. As a next step the verification of the biomarker panel would be needed to continue knowledge building, for example by conducting additional prospective studies on alcohol consumption including additional alcohol-related traditional laboratory parameters, as GGT, CDT etc. and potentially by measuring direct ethanol metabolites as EtG, PEth, FAAEs. An additional study combining GWAS and metabolomics data could provide access to the biochemical context of genetic variations, in particular when enzyme coding genes are concerned. Associations between genetic variants that associate with changes in the homeostasis of key lipids, carbohydrates, or amino acids are not only expected to display much larger effect sizes due to their direct involvement in metabolite conversion/modification, but should also provide access to the biochemical context of such variations, in particular when enzyme coding genes are concerned. This may lead to a novel approach to personalized health care based on a combination of genotyping and metabolic characterization (Gieger, Geistlinger et al. 2008). An

important aspect of the knowledge building is also the early definition of potential purpose of the candidate biomarkers. This means the early identification of the underlying intended use of the putative biomarkers and assay.

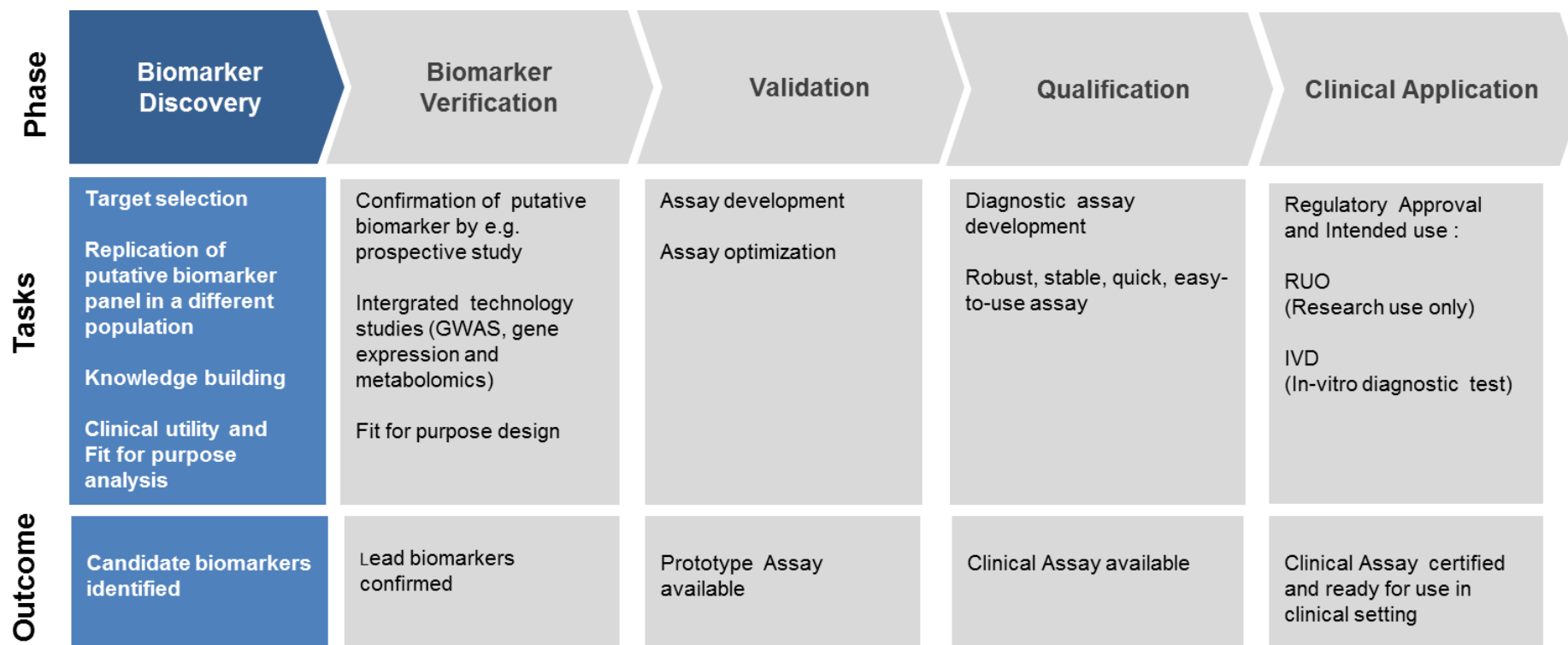


Figure 20. Laboratory diagnostic biomarker development workflow. Stages of a conventional biomarker development starting with the biomarker discovery phase and ending with the application of the biomarkers in the clinical setting. The current alcohol metabolomics study is part of the Biomarker Discovery phase. Schematic modified according to (Kitteringham, Jenkins et al. 2009).

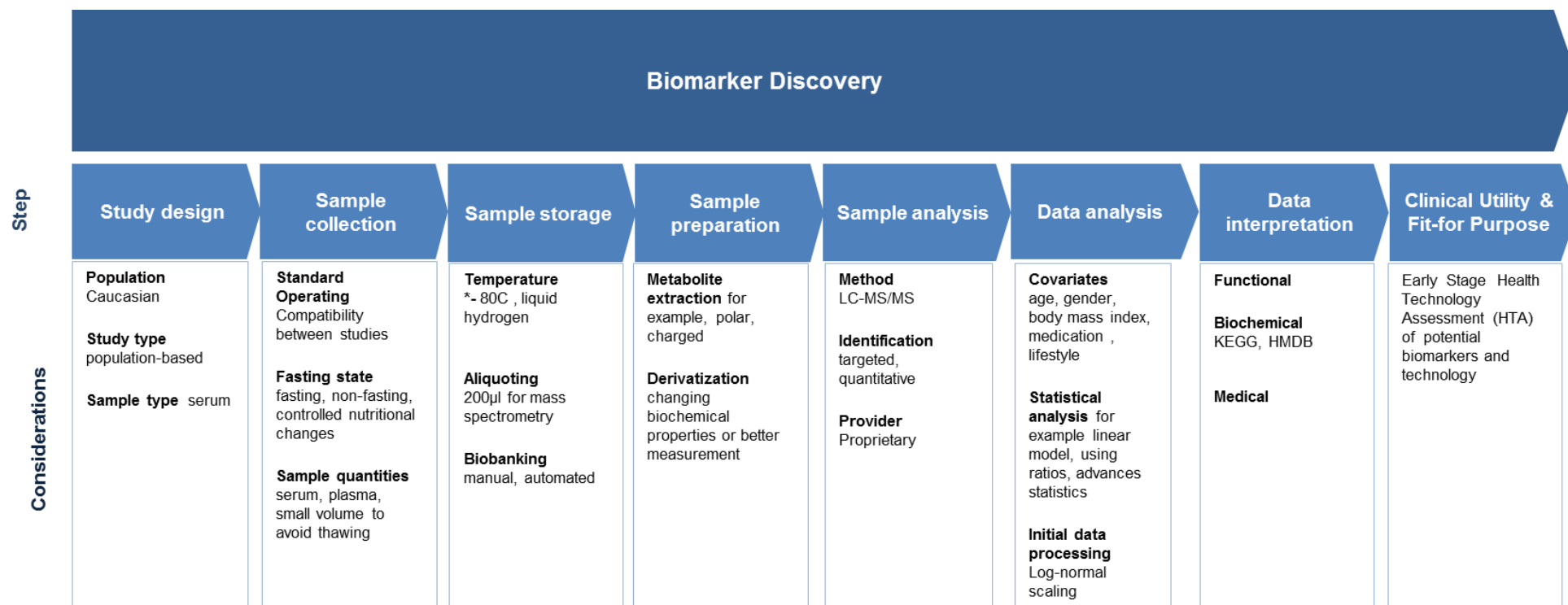


Figure 21. Key steps in alcohol biomarker discovery with metabolomics. Describes detailed key steps within the Biomarker Discovery Biomarker discovery phase with mass-spectrometry-based metabolomics. Schematic modified according to (Suhre and Gieger 2012) and <http://www.caprion.com/en/services/new-fit-for-purpose-assays.php>.

6. Conclusion and Perspectives

6.1 Metabolomics study of alcohol intake

The metabolomics study provides new insights into the impact of alcohol consumption on human metabolism. The results suggest that metabolomic profiles based on Diacylphosphatidylcholines, Lysophosphatidylcholines, Ether lipids and Sphingolipids form a new class of biomarkers for alcohol consumption. This may be of great value for the clinical assessment of alcohol intake, alcohol-specific disease detection and drug-therapy monitoring. The combined research outcomes from the current study with outcomes from other studies indicates that the identified metabolic biomarkers could also be surrogate markers for Diabetes, Huntington's disease, Ovarian endometriosis and Prostate cancer. However, further research is needed to elucidate the exact underlying mechanisms. A prospective study follow-up in large sample would help validate the predictive potential of these results.

6.2 Value of mass spectrometry for laboratory diagnostics

Due to the multi-parametric analysis capability, absolute quantification, low cost per analysis, high sensitivity, high robustness, low consumable usage, and the prototype tests making it easy for the transition from research into clinical setting, mass spectrometry is very well suited for clinical diagnostics applications and screening applications.

6.3 Perspectives in alcohol metabolomics-based research

In clinical diagnostics, multi-parameter analyses are of utmost importance for the elucidation of disease mechanisms. However, at present it is still difficult to speculate about the potential of metabolomics-based diagnostic approaches to actually improve patient's care in the future. When dedicated to diagnostic medicine,

metabolomic research might address areas where available diagnostic strategies have obvious shortcomings (see chapter 1.3). However, in an individually diagnostic work-up, it must be systematically questioned where pre-symptomatic diagnosis procedures (i.e. in alcohol-related organ damage) might be indeed helpful for the patient. If no treatment is available for a specific illness pre-symptomatic diagnosis might substantially compromise the quality of life.

6.4 Contribution of the underlying work to alcohol research

As to the current knowledge this is the first metabolomic population-based study investigating alcohol-induced metabolic changes in humans. The study bridges the knowledge gap of alcohol and its effects on the metabolic setup. The study identified alcohol-related metabolites and its potential underlying mechanisms which form an additional link/step in understanding the path between chronic moderate-to-heavy alcohol consumption and its clinical endpoints (**Figure 22**).

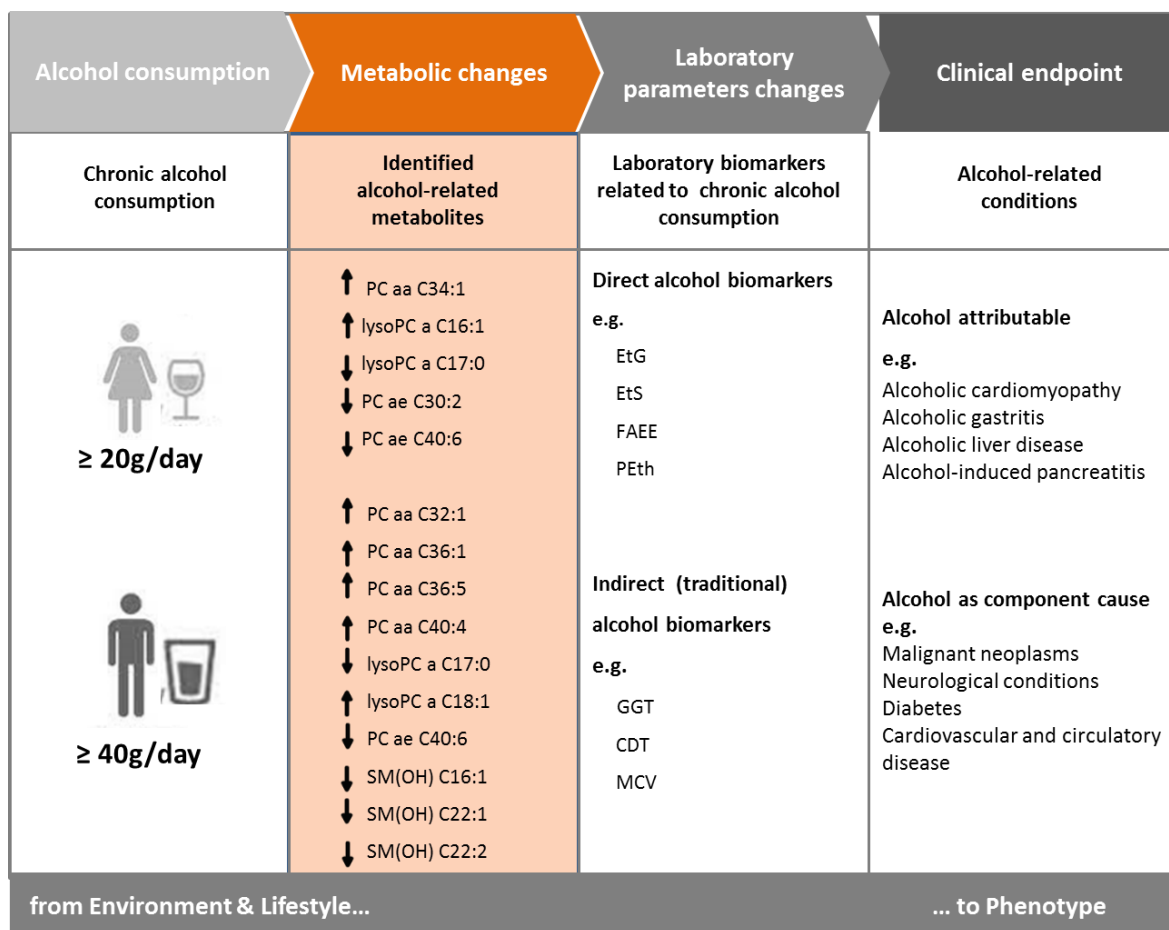


Figure 22. Alcohol metabolomics study bridges the knowledge gap between alcohol consumption and alcohol-related clinical endpoints. Orange box depicts alcohol-specific 10/5 key metabolites for males/females identified in the underlying study. The metabolites form a novel class of alcohol-related candidate biomarkers in the context of the development of alcohol-related clinical endpoints

References

(AWMF), A. d. W. M. F. (2016). S3-Leitlinie Screening, Diagnose und Behandlungalkoholbezogener Störungen.

(NCBI), N. C. f. B. I. (2010). National Library of Medicine (US), National Center for Biotechnology Information.

Agresti, A. (2002). Categorical data analysis. New York, Wiley.

Allen, J. P., P. Marques and F. Wurst (2008). "Biomarkers of alcohol use: their nature, strengths, and limitations." Mil.Med. **173**(8): v-viii.

Ashley, M. J., J. Rehm, S. Bondy, E. Single and J. Rankin (2000). "Beyond Ischemic Heart Disease: Are There other Health Benefits from Drinking Alcohol? ." Contemporary Drug Problems **27**(4): 735-777.

Bauer, D. F. (1972). "Constructing confidence sets using rank statistics." Journal of the American Statistical Association **67**: 687-690.

Baune, B. T., R. T. Mikolajczyk, G. Reymann, A. Dueterhaus, S. Fleck, H. Kratz and U. Sundermann (2005). "A 6-months assessment of the alcohol-related clinical burden at emergency rooms (ERs) in 11 acute care hospitals of an urban area in Germany." BMC Health Serv Res **5**: 73.

Bradford, B. U., T. M. O'Connell, J. Han, O. Kosyk, S. Shymonyak, P. K. Ross, J. Winnike, H. Kono and I. Rusyn (2008). "Metabolomic profiling of a modified alcohol liquid diet model for liver injury in the mouse uncovers new markers of disease." Toxicol Appl Pharmacol **232**(2): 236-243.

Breiman, L. (2001). "Random forests." Machine Learning **45** (1): 5-32.

Caro, A. A. and A. I. Cederbaum (2006). "Role of cytochrome P450 in phospholipase A2- and arachidonic acid-mediated cytotoxicity." Free Radic.Biol.Med **40**(3): 364-375.

Chace, D. H., J. C. DiPerna and E. W. Naylor (1999). "Laboratory integration and utilization of tandem mass spectrometry in neonatal screening: a model for clinical mass spectrometry in the next millennium." Acta Paediatr Suppl **88**(432): 45-47.

Conigrave, K. M., L. J. Degenhardt, J. B. Whitfield, J. B. Saunders, A. Helander, B. Tabakoff and W. I. S. Group (2002). "CDT, GGT, and AST as markers of alcohol use: the WHO/ISBRA collaborative project." Alcohol Clin Exp Res **26**(3): 332-339.

Croset, M., N. Brossard, A. Polette and M. Lagarde (2000). "Characterization of plasma unsaturated lysophosphatidylcholines in human and rat." Biochem.J **345 Pt 1**: 61-67.

Deaciuc, I. V., M. Nikolova-Karakashian, F. Fortunato, E. Y. Lee, D. B. Hill and C. J. McClain (2000). "Apoptosis and dysregulated ceramide metabolism in a murine model of alcohol-enhanced lipopolysaccharide hepatotoxicity." Alcohol Clin Exp.Res. **24**(10): 1557-1565.

Desrosiers, N. A., K. B. Scheidweiler and M. A. Huestis (2015). "Quantification of six cannabinoids and metabolites in oral fluid by liquid chromatography-tandem mass spectrometry." Drug Test Anal **7**(8): 684-694.

dhs. (2009). from <http://www.dhs.de/datenfakten/alkohol.html>.

Dimmitt, S. B., V. Rakic, I. B. Puddey, R. Baker, R. Oostryck, M. J. Adams, C. N. Chesterman, V. Burke and L. J. Beilin (1998). "The effects of alcohol on coagulation and fibrinolytic factors: a controlled trial." Blood Coagul.Fibrinolysis **9**(1): 39-45.

Dudley, R. (2014). Drunken Monkey: Why We Drink and Abuse Alcohol. California.

Eichhorst, J. C., M. L. Etter, N. Rousseaux and D. C. Lehotay (2009). "Drugs of

abuse testing by tandem mass spectrometry: a rapid, simple method to replace immunoassays." Clin Biochem **42**(15): 1531-1542.

Ewing, J. A. (1984). "Detecting alcoholism. The CAGE questionnaire." JAMA **252**(14): 1905-1907.

Fan, F., Q. Cao, C. Wang, X. Ma, C. Shen, X. W. Liu, L. P. Bu, Y. Z. Zou, K. Hu, A. J. Sun and J. B. Ge (2014). "Impact of chronic low to moderate alcohol consumption on blood lipid and heart energy profile in acetaldehyde dehydrogenase 2-deficient mice." Acta Pharmacol Sin **35**(8): 1015-1022.

Fernando, H., K. K. Bhopale, S. Kondraganti, B. S. Kaphalia and G. A. Shakeel Ansari (2011). "Lipidomic changes in rat liver after long-term exposure to ethanol." Toxicol Appl Pharmacol **255**(2): 127-137.

Fernando, H., S. Kondraganti, K. K. Bhopale, D. E. Volk, M. Neerathilingam, B. S. Kaphalia, B. A. Luxon, P. J. Boor and G. A. Shakeel Ansari (2010). "(1)H and (3)(1)P NMR lipidome of ethanol-induced fatty liver." Alcohol Clin Exp Res **34**(11): 1937-1947.

Feuerlein, W. (1994). "Alkoholismus : Ein Problem verschiedener Kulturen und Jahrhunderte." Psychiatria Danubina **6**: 55-62.

Fiehn, O. (2002). "Metabolomics--the link between genotypes and phenotypes." Plant Mol.Biol. **48**(1-2): 155-171.

Fiehn, O. (2007). <<The>> handbook of metabonomics and metabolomics. Cellular Metabolomics : The Quest for Pathway Structure. J. C. Lindon. Amsterdam, Elsevier: 561 S.

Filzmoser, P., R. Maronna and M. Werner (2008). "Outlier identification in high dimensions " Computational Statistics and Data Analysis **52** (3): 1694-1711.

Floegel, A., N. Stefan, Z. Yu, K. Muhlenbruch, D. Drogan, H. G. Joost, A. Fritsche, H. U. Haring, M. Hrabe de Angelis, A. Peters, M. Roden, C. Prehn, R. Wang-Sattler, T. Illig, M. B. Schulze, J. Adamski, H. Boeing and T. Pischon (2013). "Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach." Diabetes **62**(2): 639-648.

Floegel, A., A. Wientzek, U. Bachlechner, S. Jacobs, D. Drogan, C. Prehn, J. Adamski, J. Krumsiek, M. B. Schulze, T. Pischon and H. Boeing (2014). "Linking diet, physical activity, cardiorespiratory fitness and obesity to serum metabolite networks: findings from a population-based study." Int J Obes (Lond) **38**(11): 1388-1396.

Foti, R. S. and M. B. Fisher (2005). "Assessment of UDP-glucuronosyltransferase catalyzed formation of ethyl glucuronide in human liver microsomes and recombinant UGTs." Forensic Sci Int **153**(2-3): 109-116.

Freeman, W. M. and K. E. Vrana (2010). "Future prospects for biomarkers of alcohol consumption and alcohol-induced disorders." Alcohol Clin Exp Res **34**(6): 946-954.

Gerke, P., U. Hapke, H. J. Rumpf and U. John (1997). "Alcohol-related diseases in general hospital patients." Alcohol Alcohol **32**(2): 179-184.

Gieger, C., L. Geistlinger, E. Altmaier, M. Hrabe de Angelis, F. Kronenberg, T. Meitinger, H. W. Mewes, H. E. Wichmann, K. M. Weinberger, J. Adamski, T. Illig and K. Suhre (2008). "Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum." PLoS Genet **4**(11): e1000282.

Gika, H. G., C. Ji, G. A. Theodoridis, F. Michopoulos, N. Kaplowitz and I. D. Wilson (2012). "Investigation of chronic alcohol consumption in rodents via ultra-high-performance liquid chromatography-mass spectrometry based metabolite profiling." J Chromatogr A **1259**: 128-137.

Gika, H. G. and I. D. Wilson (2014). "Global metabolic profiling for the study of

alcohol-related disorders." Bioanalysis **6**(1): 59-77.

Goto, A., K. Sasai, S. Suzuki, T. Fukutomi, S. Ito, T. Matsushita, M. Okamoto, T. Suzuki, M. Itoh, K. Okumura-Noji and S. Yokoyama (2003). "Plasma concentrations of LPL and LCAT are in putative association with females and alcohol use which are independent negative risk factors for coronary atherosclerosis among Japanese." Clin Chim.Acta **329**(1-2): 69-76.

Griffiths, W. J., T. Koal, Y. Wang, M. Kohl, D. P. Enot and H. P. Deigner (2010). "Targeted metabolomics for biomarker discovery." Angew Chem Int Ed Engl **49**(32): 5426-5445.

Gustavsson, L. and C. Alling (1987). "Formation of phosphatidylethanol in rat brain by phospholipase D." Biochem.Biophys.Res.Commun. **142**(3): 958-963.

Haller, F., C. Prehn and J. Adamski (2010). "Quantification of steroids in human and mouse plasma using online solid phase extraction coupled to liquid chromatography tandem mass spectrometry."

Hannuksela, M. L., M. K. Liisanantti, A. E. Nissinen and M. J. Savolainen (2007). "Biochemical markers of alcoholism." Clin Chem Lab Med **45**(8): 953-961.

Hastie, T. J. and D. Pregibon (1992). Generalized linear models. Statistical Models in S. J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

Health, U. S. D. o., S. Human, Food, A. Drug, E. Center for Drug and Research (2001). Guidance for Industry - Bioanalytical Method Validation

Helander, A. (2003). "Biological markers in alcoholism." J Neural Transm Suppl(66): 15-32.

Helander, A. and O. Beck (2004). "Mass spectrometric identification of ethyl sulfate as an ethanol metabolite in humans." Clin Chem **50**(5): 936-937.

Helander, A. and Y. Zheng (2009). "Molecular species of the alcohol biomarker phosphatidylethanol in human blood measured by LC-MS." Clin Chem. **55**(7): 1395-1405.

Hillbom, M. and J. Neiman (1988). "Platelet thromboxane formation capacity after ethanol withdrawal in chronic alcoholics." Haemostasis **18**(3): 170-178.

Hoher, B. and J. Adamski (2017). "Metabolomics for clinical use and research in chronic kidney disease." Nat Rev Nephrol **13**(5): 269-284.

Holle, R., M. Happich, H. Lowel and H. E. Wichmann (2005). "KORA--a research platform for population based health research." Gesundheitswesen **67 Suppl 1**: S19-S25.

Holleboom, A. G., J. A. Kuivenhoven, M. Vergeer, G. K. Hovingh, J. N. van Miert, N. J. Wareham, J. J. Kastelein, K. T. Khaw and S. M. Boekholdt (2010). "Plasma levels of lecithin:cholesterol acyltransferase and risk of future coronary artery disease in apparently healthy men and women: a prospective case-control analysis nested in the EPIC-Norfolk population study." J Lipid Res. **51**(2): 416-421.

Hopkins, M. J., R. Sharp and G. T. Macfarlane (2001). "Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles." Gut **48**(2): 198-205.

<https://www.ncbi.nlm.nih.gov>.

Illig, T., C. Gieger, G. Zhai, W. Romisch-Margl, R. Wang-Sattler, C. Prehn, E. Altmaier, G. Kastenmuller, B. S. Kato, H. W. Mewes, T. Meitinger, M. H. de Angelis, F. Kronenberg, N. Soranzo, H. E. Wichmann, T. D. Spector, J. Adamski and K. Suhre (2010). "A genome-wide perspective of genetic variation in human metabolism 63." Nat.Genet. **42**(2): 137-141.

Iowa State University. (2017). "Mass Spectrometry Tutorial." Retrieved 01.03.2017,

from <https://www.cif.iastate.edu/mass-spec/ms-tutorial>.

Jannetto, P. J. and R. L. Fitzgerald (2016). "Effective Use of Mass Spectrometry in the Clinical Laboratory." Clinical Chemistry **62**(1): 92-98.

Jaremek, M. and K. Hiltawsky (2012). Detektion von Organismen und Bestimmung der Resistenz mit Hilfe der Massenspektrometrie. WPO. WPO. Global.

Jaremek, M., Z. Yu, M. Mangino, K. Mittelstrass, C. Prehn, P. Singmann, T. Xu, N. Dahmen, K. M. Weinberger, K. Suhre, A. Peters, A. Doring, H. Hauner, J. Adamski, T. Illig, T. D. Spector and R. Wang-Sattler (2013). "Alcohol-induced metabolomic differences in humans." Transl Psychiatry **3**: e276.

Jenkins, R. W., D. Canals and Y. A. Hannun (2009). "Roles and regulation of secretory and lysosomal acid sphingomyelinase." Cell Signal. **21**(6): 836-846.

Jenuth, J. P. (2000). "The NCBI. Publicly available tools and resources on the Web." Methods Mol Biol **132**: 301-312.

Jourdan, C., A. K. Petersen, C. Gieger, A. Doring, T. Illig, R. Wang-Sattler, C. Meisinger, A. Peters, J. Adamski, C. Prehn, K. Suhre, E. Altmaier, G. Kastenmuller, W. Romisch-Margl, F. J. Theis, J. Krumsiek, H. E. Wichmann and J. Linseisen (2012). "Body fat free mass is associated with the serum metabolite profile in a population-based study." PLoS One **7**(6): e40009.

Kaphalia, B. S., P. Cai, M. F. Khan, A. O. Okorodudu and G. A. Ansari (2004). "Fatty acid ethyl esters: markers of alcohol abuse and alcoholism." Alcohol **34**(2-3): 151-158.

Kapoor, A., K. L. Kraemer, K. J. Smith, M. S. Roberts and R. Saitz (2009). "Cost-effectiveness of screening for unhealthy alcohol use with % carbohydrate deficient transferrin: results from a literature-based decision analytic computer model." Alcohol Clin Exp Res **33**(8): 1440-1449.

Kaufman, D. W., J. P. Kelly, B. E. Wiholm, A. Laszlo, J. E. Sheehan, R. S. Koff and S. Shapiro (1999). "The risk of acute major upper gastrointestinal bleeding among users of aspirin and ibuprofen at various levels of alcohol consumption." Am.J Gastroenterol. **94**(11): 3189-3196.

Kesse, E., F. Clavel-Chapelon, N. Slimani and M. van Liere (2001). "Do eating habits differ according to alcohol consumption? Results of a study of the French cohort of the European Prospective Investigation into Cancer and Nutrition (E3N-EPIC)." Am.J Clin Nutr. **74**(3): 322-327.

Kim, S., P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang and S. H. Bryant (2016). "PubChem Substance and Compound databases." Nucleic Acids Res **44**(D1): D1202-1213.

Kitteringham, N. R., R. E. Jenkins, C. S. Lane, V. L. Elliott and B. K. Park (2009). "Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics." J Chromatogr B Analyt Technol Biomed Life Sci **877**(13): 1229-1239.

Kobayashi, M. and J. N. Kanfer (1987). "Phosphatidylethanol formation via transphosphatidylation by rat brain synaptosomal phospholipase D." J Neurochem. **48**(5): 1597-1603.

Laposata, M. (1999). "Assessment of ethanol intake. Current tests and new assays on the horizon." Am J Clin Pathol **112**(4): 443-450.

Li, S., H. Liu, Y. Jin, S. Lin, Z. Cai and Y. Jiang (2011). "Metabolomics study of alcohol-induced liver injury and hepatocellular carcinoma xenografts in mice." J Chromatogr B Analyt Technol Biomed Life Sci **879**(24): 2369-2375.

Li, Z., T. K. Hailemariam, H. Zhou, Y. Li, D. C. Duckworth, D. A. Peake, Y. Zhang, M. S. Kuo, G. Cao and X. C. Jiang (2007). "Inhibition of sphingomyelin synthase (SMS) affects intracellular sphingomyelin accumulation and plasma membrane lipid

organization." Biochim.Biophys.Acta **1771**(9): 1186-1194.

Litten, R. Z., A. M. Bradley and H. B. Moss (2010). "Alcohol biomarkers in applied settings: recent advances and future research opportunities." Alcohol Clin Exp Res **34**(6): 955-967.

Liu, C. T., R. Raghu, S. H. Lin, S. Y. Wang, C. H. Kuo, Y. J. Tseng and L. Y. Sheen (2013). "Metabolomics of ginger essential oil against alcoholic fatty liver in mice." J Agric Food Chem **61**(46): 11231-11240.

Liu, J. J., J. Y. Wang, E. Hertervig, Y. Cheng, A. Nilsson and R. D. Duan (2000). "Activation of neutral sphingomyelinase participates in ethanol-induced apoptosis in Hep G2 cells." Alcohol Alcohol **35**(6): 569-573.

Loftus, N., A. Barnes, S. Ashton, F. Michopoulos, G. Theodoridis, I. Wilson, C. Ji and N. Kaplowitz (2011). "Metabonomic investigation of liver profiles of nonpolar metabolites obtained from alcohol-dosed rats and mice using high mass accuracy MSn analysis." J Proteome Res **10**(2): 705-713.

Lowel, H., A. Doring, A. Schneider, M. Heier, B. Thorand and C. Meisinger (2005). "The MONICA Augsburg surveys--basis for prospective cohort studies." Gesundheitswesen **67 Suppl 1**: S13-S18.

Manna, S. K., A. D. Patterson, Q. Yang, K. W. Krausz, J. R. Idle, A. J. Fornace and F. J. Gonzalez (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." J Proteome Res **10**(9): 4120-4133.

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

Mastrokolas, A., R. Pool, E. Mina, K. M. Hettne, E. van Duijn, R. C. van der Mast, G. van Ommen, P. A. t Hoen, C. Prehn, J. Adamski and W. van Roon-Mom (2016). "Integration of targeted metabolomics and transcriptomics identifies deregulation of phosphatidylcholine metabolism in Huntington's disease peripheral blood samples." Metabolomics **12**: 137.

Masuo, Y., T. Imai, J. Shibato, M. Hirano, O. A. Jones, M. L. Maguire, K. Satoh, S. Kikuchi and R. Rakwal (2009). "Omic analyses unravels global molecular changes in the brain and liver of a rat model for chronic Sake (Japanese alcoholic beverage) intake." Electrophoresis **30**(8): 1259-1275.

Matsumoto, T., T. Kobayashi and K. Kamata (2007). "Role of lysophosphatidylcholine (LPC) in atherosclerosis." Curr.Med Chem. **14**(30): 3209-3220.

McShane, A. J., D. R. Bunch and S. Wang (2016). "Therapeutic drug monitoring of immunosuppressants by liquid chromatography-mass spectrometry." Clin Chim Acta **454**: 1-5.

Micheel, C. M. (2012). Evolution of Translational Omics: Lessons Learned and the Path Forward. Evolution of Translational Omics: Lessons Learned and the Path Forward. C. M. Micheel, S. J. Nass and G. S. Omenn. Washington (DC).

Mikhailidis, D. P., W. J. Jenkins, M. A. Barradas, J. Y. Jeremy and P. Dandona (1986). "Platelet function defects in chronic alcoholism." Br.Med J (Clin Res.Ed) **293**(6549): 715-718.

Millington, D. S., N. Kodo, D. L. Norwood and C. R. Roe (1990). "Tandem mass spectrometry: a new method for acylcarnitine profiling with potential for neonatal screening for inborn errors of metabolism." J Inherit Metab Dis **13**(3): 321-324.

Mittelstrass, K., J. S. Ried, Z. Yu, J. Krumsiek, C. Gieger, C. Prehn, W. Roemisch-Margl, A. Polonikov, A. Peters, F. J. Theis, T. Meitinger, F. Kronenberg, S. Weidinger, H. E. Wichmann, K. Suhre, R. Wang-Sattler, J. Adamski and T. Illig (2011). "Discovery of sexual dimorphisms in metabolic and genetic biomarkers." PLoS Genet **7**(8): e1002215.

Mittelstrass, K., J. S. Ried, Z. Yu, J. Krumsiek, C. Gieger, C. Prehn, W. Roemisch-Margl, A. Polonikov, A. Peters, F. J. Theis, T. Meitinger, F. Kronenberg, S. Weidinger, H. E. Wichmann, K. Suhre, R. Wang-Sattler, J. Adamski and T. Illig (2011). "Discovery of sexual dimorphisms in metabolic and genetic biomarkers." PLoS Genet **7**(8): e1002215.

Moayyeri, A., C. J. Hammond, A. M. Valdes and T. D. Spector (2012). "Cohort Profile: TwinsUK and Healthy Ageing Twin Study." Int J Epidemiol.

Moore, R. D., L. R. Bone, G. Geller, J. A. Mamon, E. J. Stokes and D. M. Levine (1989). "Prevalence, detection, and treatment of alcoholism in hospitalized patients." JAMA **261**(3): 403-407.

Mueller, G. C., M. F. Fleming, M. A. LeMahieu, G. S. Lybrand and K. J. Barry (1988). "Synthesis of phosphatidylethanol--a potential marker for adult males at risk for alcoholism." Proc.Natl.Acad.Sci.U.S.A **85**(24): 9778-9782.

Nagan, N. and R. A. Zoeller (2001). "Plasmalogens: biosynthesis and functions." Prog.Lipid Res. **40**(3): 199-229.

Nair, H. and W. Clarke (2017). Mass Spectrometry for the Clinical Laboratory, Elsevier Inc.

Nalesso, A., G. Viel, G. Cecchetto, D. Mioni, G. Pessa, D. Favretto and S. D. Ferrara (2011). "Quantitative profiling of phosphatidylethanol molecular species in human blood by liquid chromatography high resolution mass spectrometry." J Chromatogr.A **1218**(46): 8423-8431.

National Institute for Health and Care Excellence, N. (2011). Alcohol-use disorders: diagnosis, assessment and management of harmful drinking and alcohol dependence, NICE.

Nicholas, P. C., D. Kim, F. T. Crews and J. M. Macdonald (2008). "1H NMR-based metabolomic analysis of liver, serum, and brain following ethanol administration in rats." Chem Res Toxicol **21**(2): 408-420.

Niemela, O. (2007). "Biomarkers in alcoholism 1." Clin Chim.Acta **377**(1-2): 39-49.

NIH, N. I. o. A. A. a. A. (2006). "FOCUS ON: Biomarkers Of Fetal Alcohol Exposure And Fetal Alcohol Effects." from <https://pubs.niaaa.nih.gov/publications/arh341/56-63.htm>.

Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono and M. Kanehisa (1999). "KEGG: Kyoto Encyclopedia of Genes and Genomes." Nucleic Acids Res. **27**(1): 29-34.

Palmer, J. A., A. M. Poenitzsch, S. M. Smith, K. R. Conard, P. R. West and G. G. Cezar (2012). "Metabolic biomarkers of prenatal alcohol exposure in human embryonic stem cell-derived neural lineages." Alcohol Clin Exp Res **36**(8): 1314-1324.

Pascual, M., S. L. Valles, J. Renau-Piqueras and C. Guerri (2003). "Ceramide pathways modulate ethanol-induced cell death in astrocytes." J Neurochem. **87**(6): 1535-1545.

Popp, R., D. Malmstrom, A. G. Chambers, D. Lin, A. G. Camenzind, J. G. van der Gugten, D. T. Holmes, M. Puglia, M. Jaremek, S. Cornett, D. Suckau and C. H. Borchers (2015). "An automated assay for the clinical measurement of plasma renin activity by immuno-MALDI (iMALDI)." Biochim Biophys Acta **1854**(6): 547-558.

Pusch, W., J. H. Wurmbach, H. Thiele and M. Kostrzewa (2002). "MALDI-TOF mass

spectrometry-based SNP genotyping." Pharmacogenomics **3**(4): 537-548.

Reichel, M., E. Greiner, T. Richter-Schmidinger, O. Yedibela, P. Tripal, A. Jacobi, S. Bleich, E. Gulbins and J. Kornhuber (2010). "Increased acid sphingomyelinase activity in peripheral blood cells of acutely intoxicated patients with alcohol dependence." Alcohol Clin Exp.Res. **34**(1): 46-50.

Römisch-Margl, W., C. Prehn, R. Bogumil, C. Röhring, K. Suhre and J. Adamski (2011). "Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics." Metabolomics **8** (1): 133-142.

Royston, P. (1982). "An extension of Shapiro and Wilk's W test for normality to large samples." Applied Statistics **31** 115-124.

Ruckert, I. M., M. Heier, W. Rathmann, S. E. Baumeister, A. Doring and C. Meisinger (2011). "Association between markers of fatty liver disease and impaired glucose regulation in men and women from the general population: the KORA-F4-study." PLoS One **6**(8): e22932.

Ruf, T., G. Nagel, H. P. Altenburg, A. B. Miller and B. Thorand (2005). "Food and nutrient intake, anthropometric measurements and smoking according to alcohol consumption in the EPIC Heidelberg study." Ann Nutr Metab **49**(1): 16-25.

Saito, M., M. Saito, T. B. Cooper and C. Vadasz (2005). "Ethanol-induced changes in the content of triglycerides, ceramides, and glucosylceramides in cultured neurons." Alcohol Clin Exp.Res. **29**(8): 1374-1383.

Salem, R. O. and M. Laposata (2005). "Effects of alcohol on hemostasis." Am.J Clin Pathol. **123 Suppl**: S96-105.

Salem, R. O., M. A. Refaai, J. E. Cluette-Brown, J. W. Russo and M. Laposata (2001). "Fatty acid ethyl esters in liver and adipose tissues as postmortem markers for ethanol intake." Clin Chem. **47**(4): 722-725.

SAMHSA (2012). "The Role of Biomarkers in the Treatment of Alcohol Use Disorders, 2012 Revision." **11**(2).

Saunders, J. B., O. G. Aasland, T. F. Babor, J. R. de la Fuente and M. Grant (1993). "Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II." Addiction **88**(6): 791-804.

Schmidt, J. A., G. K. Fensom, S. Rinaldi, A. Scalbert, P. N. Appleby, D. Achaintre, A. Gicquiau, M. J. Gunter, P. Ferrari, R. Kaaks, T. Kuhn, A. Floegel, H. Boeing, A. Trichopoulou, P. Lagiou, E. Anifantis, C. Agnoli, D. Palli, M. Trevisan, R. Tumino, H. B. Bueno-de-Mesquita, A. Agudo, N. Larranaga, D. Redondo-Sanchez, A. Barricarte, J. M. Huerta, J. R. Quiros, N. Wareham, K. T. Khaw, A. Perez-Cornago, M. Johansson, A. J. Cross, K. K. Tsilidis, E. Riboli, T. J. Key and R. C. Travis (2017). "Pre-diagnostic metabolite concentrations and prostate cancer risk in 1077 cases and 1077 matched controls in the European Prospective Investigation into Cancer and Nutrition." BMC Med **15**(1): 122.

Shi, X., D. Yao and C. Chen (2012). "Identification of N-acetyltaurine as a novel metabolite of ethanol through metabolomics-guided biochemical analysis." J Biol Chem **287**(9): 6336-6349.

Shield, K. D., C. Parry and J. Rehm (2013). "Chronic Diseases and Conditions Related to Alcohol Use." Alcohol Research Current Reviews **35**(2): 155-171.

Smedman, A. E., I. B. Gustafsson, L. G. Berglund and B. O. Vessby (1999). "Pentadecanoic acid in serum as a marker for intake of milk fat: relations between intake of milk fat and metabolic risk factors." Am.J Clin Nutr. **69**(1): 22-29.

Snyder, F. (1995). "Platelet-activating factor: the biosynthetic and catabolic enzymes." Biochem.J **305 (Pt 3)**: 689-705.

Stewart, S. H., A. Reuben, W. A. Brzezinski, D. G. Koch, J. Basile, P. K. Randall and P. M. Miller (2009). "Preliminary evaluation of phosphatidylethanol and alcohol consumption in patients with liver disease and hypertension." Alcohol Alcohol **44(5)**: 464-467.

Suhre, K. and C. Gieger (2012). "Genetic variation in metabolic phenotypes: study designs and applications." Nat Rev Genet **13(11)**: 759-769.

Torrente, M. P., W. M. Freeman and K. E. Vrana (2012). "Protein biomarkers of alcohol abuse." Expert Rev Proteomics **9(4)**: 425-436.

van Buuren, S. and K. Groothuis-Oudshoorn (2012). "mice: Multivariate Imputation by Chained Equations in R." Journal of Statistical Software **45** (3).

Varga, A., P. Hansson, C. Lundqvist and C. Alling (1998). "Phosphatidylethanol in blood as a marker of ethanol consumption in healthy volunteers: comparison with other markers." Alcohol Clin Exp.Res. **22(8)**: 1832-1837.

Vazquez-Fresno, R., R. Llorach, F. Alcaro, M. A. Rodriguez, M. Vinaixa, G. Chiva-Blanch, R. Estruch, X. Correig and C. Andres-Lacueva (2012). "(1)H-NMR-based metabolomic analysis of the effect of moderate wine consumption on subjects with cardiovascular risk factors." Electrophoresis **33(15)**: 2345-2354.

Viktorov, A. V. and V. A. Yurkiv (2008). "Effects of ethanol and lipopolysaccharide on the sphingomyelin cycle in rat hepatocytes." Bull.Exp.Biol.Med. **146(6)**: 753-755.

Vogeser, M. (2003). "Liquid chromatography-tandem mass spectrometry--application in the clinical laboratory." Clin Chem Lab Med **41(2)**: 117-126.

Vogeser, M. and F. Kirchhoff (2011). "Progress in automation of LC-MS in laboratory medicine." Clin Biochem **44(1)**: 4-13.

Vogeser, M. and C. Seger (2008). "A decade of HPLC-MS/MS in the routine clinical laboratory--goals for further developments." Clin Biochem **41(9)**: 649-662.

Vogeser, M. and C. Seger (2012). "LC-MS/MS in clinical chemistry." J Chromatogr B Analyt Technol Biomed Life Sci **883-884**: 1-2.

Vouk, K., N. Hevir, M. Ribic-Pucelj, G. Haarpaintner, H. Scherb, J. Osredkar, G. Moller, C. Prehn, T. L. Rizner and J. Adamski (2012). "Discovery of phosphatidylcholines and sphingomyelins as biomarkers for ovarian endometriosis." Hum Reprod **27(10)**: 2955-2965.

Vouk, K., M. Ribic-Pucelj, J. Adamski and T. L. Rizner (2016). "Altered levels of acylcarnitines, phosphatidylcholines, and sphingomyelins in peritoneal fluid from ovarian endometriosis patients." J Steroid Biochem Mol Biol **159**: 60-69.

Wahl, S., Z. Yu, M. Kleber, P. Singmann, C. Holzapfel, Y. He, K. Mittelstrass, A. Polonikov, C. Prehn, W. Romisch-Margl, J. Adamski, K. Suhre, H. Grallert, T. Illig, R. Wang-Sattler and T. Reinehr (2012). "Childhood obesity is associated with changes in the serum metabolite profile." Obes Facts **5(5)**: 660-670.

Wang-Sattler, R., Z. Yu, C. Herder, A. C. Messias, A. Floegel, Y. He, K. Heim, M. Campillos, C. Holzapfel, B. Thorand, H. Grallert, T. Xu, E. Bader, C. Huth, K. Mittelstrass, A. Doring, C. Meisinger, C. Gieger, C. Prehn, W. Roemisch-Margl, M. Carstensen, L. Xie, H. Yamanaka-Okumura, G. Xing, U. Ceglarek, J. Thiery, G. Giani, H. Lickert, X. Lin, Y. Li, H. Boeing, H. G. Joost, M. H. de Angelis, W. Rathmann, K. Suhre, H. Prokisch, A. Peters, T. Meitinger, M. Roden, H. E. Wichmann, T. Pischon, J. Adamski and T. Illig (2012). "Novel biomarkers for pre-diabetes identified by metabolomics." Mol Syst Biol **8**: 615.

Wang-Sattler, R., Z. Yu, C. Herder, A. C. Messias, A. Floegel, Y. He, K. Heim, M. Campillos, C. Holzapfel, B. Thorand, H. Grallert, T. Xu, E. Bader, C. Huth, K.

Mittelstrass, A. Doring, C. Meisinger, C. Gieger, C. Prehn, W. Roemisch-Margl, M. Carstensen, L. Xie, H. Yamanaka-Okumura, G. Xing, U. Ceglarek, J. Thiery, G. Giani, H. Lickert, X. Lin, Y. Li, H. Boeing, H. G. Joost, M. H. de Angelis, W. Rathmann, K. Suhre, H. Prokisch, A. Peters, T. Meitinger, M. Roden, H. E. Wichmann, T. Pischon, J. Adamski and T. Illig (2012). "Novel biomarkers for pre-diabetes identified by metabolomics." Mol.Syst.Biol. **8**: 615.

Wang, J., H. Wang, H. Han, Y. Zhang, B. Yang, S. Nattel and Z. Wang (2001). "Phospholipid metabolite 1-palmitoyl-lysophosphatidylcholine enhances human ether-a-go-go-related gene (HERG) K(+) channel function." Circulation **104**(22): 2645-2648.

Wang, X., H. Lv, H. Sun, L. Liu, B. Yang, W. Sun, P. Wang, D. Zhou, L. Zhao, S. Dou, G. Zhang and H. Cao (2008). "Metabolic urinary profiling of alcohol hepatotoxicity and intervention effects of Yin Chen Hao Tang in rats using ultra-performance liquid chromatography/electrospray ionization quadruple time-of-flight mass spectrometry." J Pharm Biomed Anal **48**(4): 1161-1168.

Weiner, H., F. G. Coker and J. J. Vrbanac (1984). "Application of metabolic profiling to study the effects of ethanol on metabolism in rats." Alcohol **1**(2): 105-109.

Weltzien, H. U. (1979). "Cytolytic and membrane-perturbing properties of lysophosphatidylcholine." Biochim.Biophys.Acta **559**(2-3): 259-287.

WHO (2000). INTERNATIONAL GUIDE FOR MONITORING ALCOHOL CONSUMPTION AND RELATED HARM.

Wichmann, H. E., C. Gieger and T. Illig (2005). "KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes." Gesundheitswesen **67 Suppl 1**: S26-S30.

Wieser, A., L. Schneider, J. Jung and S. Schubert (2012). "MALDI-TOF MS in microbiological diagnostics-identification of microorganisms and beyond (mini review)." Appl Microbiol Biotechnol **93**(3): 965-974.

Wishart, D. S., T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorn Dahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert (2013). "HMDB 3.0--The Human Metabolome Database in 2013." Nucleic Acids Res **41**(Database issue): D801-807.

Wittenbecher, C., K. Muhlenbruch, J. Kroger, S. Jacobs, O. Kuxhaus, A. Floegel, A. Fritsche, T. Pischon, C. Prehn, J. Adamski, H. G. Joost, H. Boeing and M. B. Schulze (2015). "Amino acids, lipid metabolites, and ferritin as potential mediators linking red meat consumption to type 2 diabetes." Am J Clin Nutr **101**(6): 1241-1250.

Wolk, A., B. Vessby, H. Ljung and P. Barrefors (1998). "Evaluation of a biological marker of dairy fat intake." Am.J Clin Nutr. **68**(2): 291-295.

Wolk, D. M., E. J. Kaleta and V. H. Wysocki (2012). "PCR-electrospray ionization mass spectrometry: the potential to change infectious disease diagnostics in clinical and public health laboratories." J Mol Diagn **14**(4): 295-304.

World Health Organization, W. (2014). Global status report on alcohol and health.

Wu, Z. and D. L. Palmquist (1991). "Synthesis and biohydrogenation of fatty acids by ruminal microorganisms in vitro." J Dairy Sci. **74**(9): 3035-3046.

Wurst, F. M., S. Alexson, M. Wolfersdorf, G. Bechtel, S. Forster, C. Alling, S. Aradottir, K. Jachau, P. Huber, J. P. Allen, V. Auwarter and F. Pragst (2004). "Concentration of fatty acid ethyl esters in hair of alcoholics: comparison to other biological state markers and self reported-ethanol intake." Alcohol Alcohol **39**(1): 33-38.

www.statisticssolutions.com. "Correlation (Pearson, Kendall, Spearman)."

Xia, J., R. Mandal, I. V. Sinelnikov, D. Broadhurst and D. S. Wishart (2012). "MetaboAnalyst 2.0--a comprehensive server for metabolomic data analysis." *Nucleic Acids Res* **40**(Web Server issue): W127-133.

Xu, T., C. Holzapfel, X. Dong, E. Bader, Z. Yu, C. Prehn, K. Perstorfer, M. Jaremek, W. Roemisch-Margl, W. Rathmann, Y. Li, H. E. Wichmann, H. Wallaschofski, K. H. Ladwig, F. Theis, K. Suhre, J. Adamski, T. Illig, A. Peters and R. Wang-Sattler (2013). "Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study." *BMC Med* **11**: 60.

Yoseph, B. P., E. Breed, C. E. Overgaard, C. J. Ward, Z. Liang, M. E. Wagener, D. R. Lexcen, E. R. Luszczek, G. J. Beilman, E. M. Burd, A. B. Farris, D. M. Guidot, M. Koval, M. L. Ford and C. M. Coopersmith (2013). "Chronic alcohol ingestion increases mortality and organ injury in a murine model of septic peritonitis." *PLoS One* **8**(5): e62792.

Yu, Z., G. Kastenmuller, Y. He, P. Belcredi, G. Moller, C. Prehn, J. Mendes, S. Wahl, W. Roemisch-Margl, U. Ceglarek, A. Polonikov, N. Dahmen, H. Prokisch, L. Xie, Y. Li, H. E. Wichmann, A. Peters, F. Kronenberg, K. Suhre, J. Adamski, T. Illig and R. Wang-Sattler (2011). "Differences between human plasma and serum metabolite profiles 2." *PLoS One* **6**(7): e21230.

Yu, Z., G. Zhai, P. Singmann, Y. He, T. Xu, T. D. Spector, T. Illig and R. Wang-Sattler (2012). "Human serum metabolic profiles are age dependent." *Aging Cell*.

List of Figures

Figure 1. Main short-term and long-term effects of alcohol consumption..	10
Figure 2. Current traditional alcohol biomarkers are limited.....	13
Figure 3. Mass spectrometry technology principle.....	16
Figure 4. Common technologies and configurations of mass spectrometry instrumentation.....	17
Figure 5. Selection process of study participants.	26
Figure 6. Selection path of key metabolites.....	33
Figure 7. Global analysis in males shows significant metabolic changes only between moderate-to-heavy drinkers (MHD) and low-drinkers (LD).....	40
Figure 8. Global analysis in females shows significant metabolic change only between moderate-to-heavy drinkers (MHD) and low drinkers (LD).....	45
Figure 9. Alcohol-specific metabolite clusters and key metabolites.....	49
Figure 10. Alcohol-specific key metabolite panel gives a good-to-fair prediction for alcohol consumption.....	51
Figure 11. Non-linear relationship between alcohol consumption and metabolite concentrations in males.....	56
Figure 12. Non-linear relationship between alcohol consumption and metabolite concentrations in females.....	57
Figure 13. Correlation between 10/5 key metabolites in males/females and laboratory parameters HDL and triglycerides	59
Figure 14. Summary of pathway analysis: Moderate-to-heavy alcohol consumption has the biggest impact on the Glycerophospholipid metabolism, Ether lipid metabolism and Sphingolipid metabolism.	61

Figure 15. Potential alcohol-induced biochemical mechanisms.	67
Figure 16. Mass spectrometry is used in almost all areas of laboratory medicine. ...	77
Figure 17. Mass spectrometry methods cover a wide spectrum of different analytes used in laboratory diagnostics.	78
Figure 18. Mass spectrometry applications still require significant hands-on-time due to manual sample preparation.	81
Figure 19. Current clinical laboratory processes with mass spectrometry still have a relatively long Turn-around-Time (TAT)	82
Figure 20. Laboratory diagnostic biomarker development workflow.	85
Figure 21. Key steps in alcohol biomarker discovery with metabolomics.	86
Figure 22. Alcohol metabolomics study bridges the knowledge gap between alcohol consumption and alcohol-related clinical endpoints.	89

List of Tables

Table 1. Current studies on alcohol-induced metabolic perturbations..	20
Table 2. Drinkers - study population characteristics of KORA F4 discovery, KORA F3 and TwinsUK replication data sets.	38
Table 3. Abstainers - study population characteristics of KORA F4 discovery, KORA F3 and TwinsUK replication data sets.	39
Table 4. Result of logistic regression model analysis in KORA F4 in males	43
Table 5. Result of logistic regression model analysis in KORA F4 females..	47
Table 6. Result of logistic regression analysis of alcohol-specific key metabolites in males.	53
Table 7. Result of logistic regression analysis of alcohol-specific key metabolites in females.	54
Table 8. Detailed results from Pathway analysis..	61
Table 9. Alcohol-related key metabolites are also associated with other phenotypes and diseases states.....	63

Publications

Main publication underlying the present thesis:

Jaremek M, Yu Z, Mangino M, Mittelstrass K, Prehn C, Singmann P, Xu T, Dahmen N, Weinberger KM, Suhre K, Peters A, Doering A, Hauner H, Adamski J, Illig T, Spector TD, Wang- Sattler R: Alcohol-induced metabolomic differences in humans. *Transl Psychiatry*. 2013 Jul 2;3: e276.

Further work in the field of mass spectrometry-based biomarker discovery

Xu T, Holzapfel C, Dong X, Bader E, Yu Z, Prehn C, Perstorfer K, **Jaremek M**, Roemisch-Margl W, Rathmann W, Li Y, Wichmann HE, Wallaschofski H, Ladwig KH, Theis F, Suhre K, Adamski J, Illig T, Peters A, Wang-Sattler R.: Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med*. 2013 Mar 4; 11:60. doi: 10.1186/1741-7015-11-60.

Popp, R. Malmstrom, D. Chambers, A. G. Lin, D. Camenzind, A. G. van der Gugten, J. G. Holmes, D. T. Pugia, **M. Jaremek**, M. Cornett, S. Suckau, D. Borchers, C. H.: An automated assay for the clinical measurement of plasma renin activity by immuno-MALDI (iMALDI). *Biochim Biophys Acta*. 2015 Jun. 1854; 6; 547-58; doi: 10.1016/j.bbapap.2014.10.008

Kerstin Wiesner, **Marta Jaremek**, Roland Pohle, Oliver von Sicard, Evamaria Stuetz: Monitoring of Bacterial Growth and Rapid Evaluation of Antibiotic Susceptibility by Headspace Gas Analysis. Poster. EUROSENSORS 2014, the 28th European Conference on Solid-State Transducers. *Procedia Engineering*. Volume 87, 2014, Pages 332–335. doi: 10.1016/j.proeng.2014.11.750.

Marta Jaremek und Karsten Hiltawsky

Detektion von Organismen und Bestimmung der Resistenz mit Hilfe der Massenspektrometrie. International Patent WO2012123262 A1, issued September 20, 2012

Frank Maixner, Thorsten Overath, Dennis Linke, Marek Janko, Gea Guerriero, Bart H J van den Berg, Bjoern Stade, Petra Leidinger, Christina Backes, **Marta Jaremek**, Benny Kneissl, Benjamin Meder, Andre Franke, Eduard Egarter-Vigl, Eckart Meese, Andreas Schwarz, Andreas Tholey, Albert Zink, Andreas Keller. Paleoproteomic study of the Iceman's brain tissue. *Cell Mol Life Sci* 2013 Oct 6;70(19):3709-22. Epub 2013 Jun 6.

Danksagung

Meine Doktorarbeit entstand in der Zeit von 2010 bis 2015 am Helmholtz Zentrum München. Ich möchte mich insbesondere bei Prof. Jerzy Adamski, Prof. Hans Hauner und Prof. Thomas Illig und für die Möglichkeit danken meine Doktorarbeit an ihren Instituten durchzuführen, sowie für den gewährten wissenschaftlichen und zeitlichen Gestaltungsfreiraum.

Mein besonderer Dank geht an Frau Dr. Rui Wang-Sattler für die fachliche Unterstützung.

Ebenfalls möchte ich danken den Kollegen vom Helmholtz Institut für Epidemiologie und Experimentelle Genetik für die sehr gute Zusammenarbeit der letzten Jahre.