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

Marta Jaremek


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

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.
<table>
<thead>
<tr>
<th>Short-term effects</th>
<th>Long-term effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevated transaminases, elevated GGT</td>
<td>Fatty liver, Alcoholic hepatitis (ASH), Liver cirrhosis, Hepatocellular Carcinoma, Hepatocellular carcinoma</td>
</tr>
<tr>
<td>Acute Pancreatitis</td>
<td>Chronic pancreatitis, Pancreatic cancer, Diabetes mellitus</td>
</tr>
<tr>
<td>Arterial hypertension</td>
<td>Cardiomyopathy, Arrhythmia</td>
</tr>
<tr>
<td>Gastritis, astroesophageal reflux disease</td>
<td>Esophageal Varices, Mallory weiss syndrome, Esophageal cancer</td>
</tr>
<tr>
<td>Headache, Peripheral neuropathy</td>
<td>Alcohol withdrawal syndrome, Wernicke encephalopathy, Dementia, Cerebral atrophy, Peripheral neuropathy</td>
</tr>
<tr>
<td>Fetal alcohol spectrum disorders (FASDs)</td>
<td>Sexual dysfunction, Amenorrhoe, Early menopause</td>
</tr>
<tr>
<td>Lower bone mineral density, Fracture</td>
<td>Fracture</td>
</tr>
</tbody>
</table>

**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

**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 “Oomics” 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 (Hocher 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 H\(^1\)-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,

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

*Table 1. Current studies on alcohol-induced metabolic perturbations.* Shows the result of literature baseline analysis using PubMed.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Phenotype/Disease state under study</th>
<th>Outcome</th>
<th>Metabolomics-based method</th>
<th>Sample (for metabolomics analysis)</th>
<th>Technology</th>
<th>Instrumentation configuration</th>
<th>Candidate Biomarkers identified/Significant metabolite changes (as described in the publication)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manna et al 2010</td>
<td>Alcohol-induced Liver Disease (ALD)</td>
<td>Demonstration of metabolomics approach to identify early, noninvasive biomarkers of ALD pathogenesis</td>
<td>metabolomics</td>
<td>urine</td>
<td>MS</td>
<td>UPLC-ESI-QTOF-MS</td>
<td>Ethyl-sulfate, Ethyl-β-D-glucuronide, 4-hydroxyphenylacetic acid, 4-hydroxyphenylacetic acid sulfate, 2-hydroxyphenylacetic acid, adipic acid, pimelic acid, indole-3-lactic acid</td>
</tr>
<tr>
<td>Fernando et al 2011</td>
<td>Alcohol Liver Disease (ALD)</td>
<td>n.a.</td>
<td>Lipidomics</td>
<td>plasma/liver tissue</td>
<td>NMR</td>
<td>1H-NMR 31P-NMR</td>
<td>Phosphatidylcholine, Lysophosphatidylcholine, Phosphatidylinerine, Phosphatidylethanolamine, Lyso-Phosphatidylethanolamine</td>
</tr>
<tr>
<td>Li et al 2011</td>
<td>Liver injury Hepatocellular carcinoma (HCC)</td>
<td>LPC profiles in serum may be biomarkers for liver injury and hepatocellular carcinoma</td>
<td>Metabolomics</td>
<td>serum</td>
<td>MS</td>
<td>UPLC/Q-TOF-MS MS</td>
<td>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)</td>
</tr>
</tbody>
</table>

*Table 1 (continued).* Current studies on alcohol-induced metabolic perturbations. Shows the result of literature baseline analysis using PubMed.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Phenotype/Disease state under study</th>
<th>Outcome</th>
<th>Metabolomics-based method</th>
<th>Sample (for metabolomics analysis)</th>
<th>Technology</th>
<th>Instrumentation configuration</th>
<th>Candidate Biomarkers identified/Significant metabolite changes (as described in the publication)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loftus et al 2011</td>
<td>Alcoholism</td>
<td>Increased metabolite levels in livers in rat and mouse by alcohol treatment.</td>
<td>Metabolite profiling</td>
<td>liver tissue</td>
<td>MS</td>
<td>LCMS-IT-TOF</td>
<td>Fatty acyl 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</td>
</tr>
<tr>
<td>Manna et al 2011</td>
<td>Alcohol-induced Liver Disease (ALD)</td>
<td>Indole-3-lactic Acid and Phenyllactic Acid are potential candidates for conserved and pathology-specific for early stages of ALD</td>
<td>metabolomics</td>
<td>urine</td>
<td>MS</td>
<td>UPLC-ESIQTOF-MS</td>
<td>Indole-3-lactic Acid Phenyllactic Acid</td>
</tr>
<tr>
<td>Gika et al 2012</td>
<td>Relationship between chronic alcohol consumption and metabolic profiles</td>
<td>A large number of metabolites were seen to differ between control and alcohol-treated animals, for both biofluids.</td>
<td>Metabolic profiling</td>
<td>plasma, urine</td>
<td>MS</td>
<td>UHPLC-TOF MS</td>
<td>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)</td>
</tr>
<tr>
<td>Palmer et al 2012</td>
<td>Relationship between ethanol exposure and metabolite levels of human embryonic stem cells</td>
<td>Ethanol exposure induces changes to metabolite profile of human embryoid bodies</td>
<td>Metabolomics</td>
<td>human embryonic stem (hES) cells</td>
<td>MS</td>
<td>ESI-QTOF-MS</td>
<td>L-thyroxine S'-methylthioadenosine tryptophan metabolites L-kyurenine indoleacetaldehyde</td>
</tr>
<tr>
<td>Shi et al 2012</td>
<td>Relationship between Ethanol and Hepatic Steatosis</td>
<td>Metabolomic analysis of urinary metabolites revealed time- and dose-dependent changes in the chemical composition of urine.</td>
<td>Metabolomics</td>
<td>urine</td>
<td>MS</td>
<td>ESI-QTOF-MS</td>
<td>4-Hydroxyphenylacetic acid sulfate; Ethyl glucuronide (EtG); Ethyl sulfate (EtS) N-Acetyleneuraminic acid</td>
</tr>
</tbody>
</table>

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

**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 ≥ 20g/day and males ≥ 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, Kastenmüller 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.

![Diagram](image.png)

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 (rho) 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 HMBD 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 (Jenuth 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).
Table 2. Drinkers - study population characteristics of KORA F4 discovery, KORA F3 and TwinsUK replication data sets.
Drinking status defined as a alcohol consumption < 40g/day in males, < 20 g/day in females. b alcohol 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. 629 participants were available for replication analysis, for 277 participants HDL-C and triglyceride data was available, c Mann-Whitney test (two-sided), d Fisher’s exact test (two-sided), significance level α= 0.05. Significant p-values are in bold.
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.

<table>
<thead>
<tr>
<th></th>
<th>Discovery KORA F4 (n=728)</th>
<th>Replication KORA F3 (n= 109)</th>
<th>Replication TwinsUK (n=99)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>men (n=233)</td>
<td>women (n=495)</td>
<td>men (n=37)</td>
</tr>
<tr>
<td>participants n(%)</td>
<td>233 (32)</td>
<td>495 (68)</td>
<td>37 (33.94)</td>
</tr>
<tr>
<td>age(years) mean±SD</td>
<td>53.7 ± 14.0</td>
<td>54.5 ± 13.2</td>
<td>65.4 ± 5.8</td>
</tr>
<tr>
<td>BMI (kg/m²) mean±SD</td>
<td>27.7 ± 4.6</td>
<td>27.7 ± 5.4</td>
<td>28.0 ± 3.8</td>
</tr>
<tr>
<td>smoker n(%)</td>
<td>61 (26.180)</td>
<td>86 (17.37)</td>
<td>6 (16.21622)</td>
</tr>
<tr>
<td>HDL(mg/dl) mean±SD</td>
<td>45.8 ± 11.1</td>
<td>58.4 ± 14.1</td>
<td>48.4 ± 10.3</td>
</tr>
<tr>
<td>triglyceride (mg/dl)</td>
<td>141.5 ± 134.2</td>
<td>105.4 ± 55.3</td>
<td>189.7 ± 142.0</td>
</tr>
</tbody>
</table>

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).

![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 (*)](image-url)

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 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 asterix (*).
### 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.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Abstainer Mean ± SD (µM)</th>
<th>LD Mean ± SD (µM)</th>
<th>MHD Mean ± SD (µM)</th>
<th>Odds ratio (95% CI)</th>
<th>P-value</th>
<th>Odds ratio (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C16:1</td>
<td>0.0362 ± 0.0104</td>
<td>0.0372±0.0109</td>
<td>0.0419±0.0128</td>
<td>0.99</td>
<td>9.33E-01</td>
<td>1.34</td>
<td>1.43E-04</td>
</tr>
<tr>
<td>PC aa C30:0</td>
<td>4.23±1.34</td>
<td>4.35±1.29</td>
<td>5.06±1.55</td>
<td>1.01</td>
<td>9.39E-01</td>
<td>1.37</td>
<td>4.19E-05</td>
</tr>
<tr>
<td>PC aa C32:0</td>
<td>14.40±3.20</td>
<td>14.67±2.82</td>
<td>16.44±3.59</td>
<td>1.04</td>
<td>8.26E-01</td>
<td>1.40</td>
<td>3.21E-05</td>
</tr>
<tr>
<td>PC aa C32:1</td>
<td>17.06±7.75</td>
<td>19.27±9.16</td>
<td>31.04±16.87</td>
<td>1.29</td>
<td>2.64E-01</td>
<td>2.34</td>
<td>1.15E-18</td>
</tr>
<tr>
<td>PC aa C32:3</td>
<td>0.441±0.1054</td>
<td>0.434±0.0918</td>
<td>0.424±0.0955</td>
<td>1.16</td>
<td>3.75E-01</td>
<td>0.68</td>
<td>3.20E-06</td>
</tr>
<tr>
<td>PC aa C34:1</td>
<td>223.33±49.97</td>
<td>232.38±48.19</td>
<td>276.21±67.34</td>
<td>1.08</td>
<td>6.74E-01</td>
<td>1.96</td>
<td>9.25E-13</td>
</tr>
<tr>
<td>PC aa C34:4</td>
<td>1.88±0.64</td>
<td>2.08±0.72</td>
<td>2.49±0.83</td>
<td>1.67</td>
<td>1.05E-02</td>
<td>1.41</td>
<td>1.80E-05</td>
</tr>
<tr>
<td>PC aa C36:1</td>
<td>51.15±14.11</td>
<td>51.35±11.89</td>
<td>59.70±16.38</td>
<td>0.80</td>
<td>1.80E-01</td>
<td>1.48</td>
<td>6.63E-06</td>
</tr>
<tr>
<td>PC aa C36:4</td>
<td>199.44±46.24</td>
<td>215.33±49.60</td>
<td>245.90±54.86</td>
<td>1.61</td>
<td>1.62E-02</td>
<td>1.49</td>
<td>2.10E-06</td>
</tr>
<tr>
<td>PC aa C36:5</td>
<td>24.35±11.46</td>
<td>28.69±12.93</td>
<td>36.59±17.55</td>
<td>2.08</td>
<td>5.75E-03</td>
<td>1.46</td>
<td>3.98E-07</td>
</tr>
<tr>
<td>PC aa C38:0</td>
<td>3.12±0.84</td>
<td>3.23±0.91</td>
<td>3.03±0.81</td>
<td>1.00</td>
<td>9.94E-01</td>
<td>0.68</td>
<td>4.48E-06</td>
</tr>
<tr>
<td>PC aa C38:5</td>
<td>56.14±12.69</td>
<td>60.34±14.04</td>
<td>69.93±18.12</td>
<td>1.73</td>
<td>1.36E-02</td>
<td>1.55</td>
<td>2.10E-07</td>
</tr>
<tr>
<td>PC aa C40:4</td>
<td>3.98±1.17</td>
<td>4.04±1.17</td>
<td>4.85±1.59</td>
<td>1.30</td>
<td>1.55E-01</td>
<td>1.57</td>
<td>1.25E-07</td>
</tr>
<tr>
<td>PC aa C40:5</td>
<td>11.02±2.99</td>
<td>11.21±3.00</td>
<td>13.00±4.16</td>
<td>1.37</td>
<td>1.10E-01</td>
<td>1.45</td>
<td>1.86E-05</td>
</tr>
<tr>
<td>PC aa C42:0</td>
<td>0.56±0.16</td>
<td>0.57±0.18</td>
<td>0.52±0.15</td>
<td>0.89</td>
<td>5.24E-01</td>
<td>0.67</td>
<td>5.34E-06</td>
</tr>
<tr>
<td>PC aa C42:1</td>
<td>0.28±0.08</td>
<td>0.29±0.08</td>
<td>0.27±0.08</td>
<td>1.05</td>
<td>7.59E-01</td>
<td>0.73</td>
<td>2.18E-04</td>
</tr>
<tr>
<td>PC ae C30:2</td>
<td>0.65±0.15</td>
<td>0.15±0.04</td>
<td>0.14±0.04</td>
<td>1.08</td>
<td>6.65E-01</td>
<td>0.73</td>
<td>7.56E-05</td>
</tr>
<tr>
<td>PC ae C34:2</td>
<td>12.07±3.06</td>
<td>12.01±3.08</td>
<td>11.13±2.83</td>
<td>0.86</td>
<td>3.77E-01</td>
<td>0.56</td>
<td>6.04E-11</td>
</tr>
<tr>
<td>PC ae C34:3</td>
<td>7.73±2.31</td>
<td>7.93±2.28</td>
<td>7.73±2.22</td>
<td>1.17</td>
<td>4.18E-01</td>
<td>0.65</td>
<td>2.95E-06</td>
</tr>
<tr>
<td>PC ae C36:1</td>
<td>7.85±1.87</td>
<td>7.85±1.65</td>
<td>7.69±1.77</td>
<td>0.87</td>
<td>4.43E-01</td>
<td>0.73</td>
<td>9.89E-05</td>
</tr>
</tbody>
</table>
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 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 asterix (*).
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 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 asterix (*).
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 dendogram 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.
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.
<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Discovery KORA F4 (n=1144)</th>
<th>Replication KORA F3 (n=154)</th>
<th>Meta-analysis Discovery + Replication fixed effects (n=1298)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LD mean±SD (µM) (^a)</td>
<td>MHD mean±SD (µM) (^a)</td>
<td>P-value(^{b,c})</td>
</tr>
<tr>
<td>PC aa C32:1</td>
<td>19.3±9.2</td>
<td>31.0±16.9</td>
<td>1.1E-18</td>
</tr>
<tr>
<td>PC aa C36:5</td>
<td>28.7±12.9</td>
<td>36.6±17.6</td>
<td>3.9E-07</td>
</tr>
<tr>
<td>PC aa C40:4</td>
<td>4.0±1.2</td>
<td>4.9±1.6</td>
<td>1.2E-07</td>
</tr>
<tr>
<td>PC aa C36:1</td>
<td>51.4±11.9</td>
<td>59.7±16.4</td>
<td>6.6E-06</td>
</tr>
<tr>
<td>lysoPC a C17:0</td>
<td>1.7±0.5</td>
<td>1.5±0.5</td>
<td>5.4E-10</td>
</tr>
<tr>
<td>lysoPC a C18:1</td>
<td>20.0±5.5</td>
<td>22.7±7.4</td>
<td>1.6E-05</td>
</tr>
<tr>
<td>PC ae C40:6</td>
<td>4.9±1.2</td>
<td>4.4±1.1</td>
<td>4.3E-10</td>
</tr>
<tr>
<td>SM(OH) C16:1</td>
<td>3.1±0.7</td>
<td>2.7±0.7</td>
<td>6.8E-12</td>
</tr>
<tr>
<td>SM(OH) C22:1</td>
<td>12.6±2.5</td>
<td>11.4±2.9</td>
<td>3.4E-09</td>
</tr>
<tr>
<td>SM(OH) C22:2</td>
<td>10.2±2.1</td>
<td>9.1±2.3</td>
<td>4.0E-13</td>
</tr>
<tr>
<td></td>
<td>LD mean±SD (µM) (^a)</td>
<td>MHD mean±SD (µM) (^a)</td>
<td>P-value(^{b,d})</td>
</tr>
<tr>
<td></td>
<td>20.7±9.7</td>
<td>35.0±16.5</td>
<td>7.3E-05</td>
</tr>
<tr>
<td></td>
<td>31.8±19.8</td>
<td>44.0±278</td>
<td>4.9E-02</td>
</tr>
<tr>
<td></td>
<td>3.7±1.0</td>
<td>4.7±1.5</td>
<td>1.6E-02</td>
</tr>
<tr>
<td></td>
<td>53.0±11.2</td>
<td>66.6±15.9</td>
<td>9.9E-04</td>
</tr>
<tr>
<td></td>
<td>1.6±0.4</td>
<td>1.3±0.4</td>
<td>3.4E-02</td>
</tr>
<tr>
<td></td>
<td>20.9±5.0</td>
<td>5.0±1.1</td>
<td>3.5E-02</td>
</tr>
<tr>
<td></td>
<td>2.7±0.7</td>
<td>2.4±0.6</td>
<td>2.0E-03</td>
</tr>
<tr>
<td></td>
<td>10.0±2.5</td>
<td>9.3±2.1</td>
<td>1.2E-02</td>
</tr>
<tr>
<td></td>
<td>8.1±2.1</td>
<td>7.7±1.7</td>
<td>2.3E-02</td>
</tr>
<tr>
<td></td>
<td>31.6±16.8</td>
<td>37.7±19.5</td>
<td>8.6E-23</td>
</tr>
<tr>
<td></td>
<td>29.0±13.9</td>
<td>4.8±1.6</td>
<td>5.6E-08</td>
</tr>
<tr>
<td></td>
<td>4.0±1.2</td>
<td>4.8±1.6</td>
<td>2.8E-09</td>
</tr>
<tr>
<td></td>
<td>51.5±11.8</td>
<td>60.7±16.5</td>
<td>9.9E-09</td>
</tr>
<tr>
<td></td>
<td>1.7±0.5</td>
<td>1.5±0.5</td>
<td>4.1E-11</td>
</tr>
<tr>
<td></td>
<td>19.9±5.4</td>
<td>22.5±7.1</td>
<td>3.9E-06</td>
</tr>
<tr>
<td></td>
<td>4.9±1.2</td>
<td>4.5±1.1</td>
<td>1.5E-10</td>
</tr>
<tr>
<td></td>
<td>3.0±0.7</td>
<td>2.6±0.7</td>
<td>1.2E-13</td>
</tr>
<tr>
<td></td>
<td>12.3±2.6</td>
<td>11.1±2.9</td>
<td>4.4E-10</td>
</tr>
<tr>
<td></td>
<td>9.9±2.2</td>
<td>8.9±2.3</td>
<td>1.3E-13</td>
</tr>
</tbody>
</table>

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.
**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, 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 < 0.01 females (Bonferroni corrected), e alcohol consumption < 40g/day males, < 20g/day females, f alcohol consumption ≥ 40g/day males, ≥ 20g/day females, g meta-analysis consist of KORA F4 discovery, KORA F3 and TwinsUK replication data sets, h logistic regression analysis adjusted for age, BMI, smoking with n=629 study participants. Significant P-values are represented in **bold**.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Discovery KORA F4 (n=946)</th>
<th>Replication KORA F3 (n=107)</th>
<th>Meta-analysis Discovery + Replication fixed effect (n=1053)</th>
<th>Replication TwinsUK (n=277)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LD mean±SD (µM)(^a)</td>
<td>MHD mean±SD (µM)(^a)</td>
<td>P-value(^b,c)</td>
<td>LD mean±SD (µM)(^a)</td>
</tr>
<tr>
<td>PC aa C34:1</td>
<td>241.1±52.9</td>
<td>259.4±56.3</td>
<td><strong>1.0E-04</strong></td>
<td>274.8±72.8</td>
</tr>
<tr>
<td>lysoPC a C16:1</td>
<td>2.8±0.9</td>
<td>3.1±1.0</td>
<td><strong>4.7E-05</strong></td>
<td>2.5±1.0</td>
</tr>
<tr>
<td>lysoPC a C17:0</td>
<td>1.8±0.5</td>
<td>1.6±0.5</td>
<td><strong>2.8E-04</strong></td>
<td>1.6±0.5</td>
</tr>
<tr>
<td>PC ae C30:2</td>
<td>0.17±0.04</td>
<td>0.16±0.04</td>
<td><strong>4.0E-05</strong></td>
<td>0.17±0.03</td>
</tr>
<tr>
<td>PC ae C40:6</td>
<td>5.4±1.3</td>
<td>5.0±1.2</td>
<td><strong>2.4E-07</strong></td>
<td>6.1±1.5</td>
</tr>
</tbody>
</table>

**Note:** The table includes results of logistic regression analysis for alcohol-specific key metabolites in females, with data from Discovery KORA F4, Replication KORA F3, and TwinsUK replication sets. 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 the y-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.

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) (Mastrokolias, 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).
### 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.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Gender</th>
<th>Alcohol consumption</th>
<th>Smoking</th>
<th>Red meat consumption</th>
<th>Obesity</th>
<th>Diabetes</th>
<th>Ovarian endometriosis</th>
<th>Huntington’s disease</th>
<th>Prostate cancer</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC aa C36:5</td>
<td>↑*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Jaremek, Yu et al. 2013) (Mastrokolics, Pool et al. 2016)</td>
</tr>
<tr>
<td>lysoPC a C16:1</td>
<td>↑**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Jaremek, Yu et al. 2013)</td>
</tr>
<tr>
<td>SM (OH) C16:1</td>
<td>↓*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Jaremek, Yu et al. 2013)</td>
</tr>
<tr>
<td>SM (OH) C22:1</td>
<td>↓*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Jaremek, Yu et al. 2013)</td>
</tr>
<tr>
<td>PC aa C34:1</td>
<td>↑**</td>
<td>↑*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Jaremek, Yu et al. 2013) (Xu, Holzapfel et al. 2013)</td>
</tr>
<tr>
<td>PC aa C32:1</td>
<td>↑*</td>
<td>↑***</td>
<td>↑</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Jaremek, Yu et al. 2013) (Xu, Holzapfel et al. 2013) (Floegel, Stefan et al. 2013)</td>
</tr>
<tr>
<td>PC aa C36:1</td>
<td>↑*</td>
<td>↑*</td>
<td>↑</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Jaremek, Yu et al. 2013) (Xu, Holzapfel et al. 2013) (Floegel, Stefan et al. 2013)</td>
</tr>
<tr>
<td>PC ae C40:6</td>
<td>↓*</td>
<td>↓**</td>
<td>↓*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Jaremek, Yu et al. 2013) (Xu, Holzapfel et al. 2013) (Floegel, Stefan et al. 2013)</td>
</tr>
<tr>
<td>lysoPC a C17:0</td>
<td>↓*</td>
<td>↓**</td>
<td>↓*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Jaremek, Yu et al. 2013) (Wittenbecher, Muhlenbruch et al. 2015)</td>
</tr>
<tr>
<td>lysoPC a C18:0</td>
<td>↑*</td>
<td>↑*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Jaremek, Yu et al. 2013) (Mittelstrass, Ried et al. 2011) (Wahl, Yu et al. 2012)</td>
</tr>
<tr>
<td>PC ae C30:2</td>
<td>↑**</td>
<td>↓**</td>
<td>↑</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Jaremek, Yu et al. 2013) (Mittelstrass, Ried et al. 2011) (Schmidt, Fensom et al. 2017)</td>
</tr>
</tbody>
</table>


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.

**Sphingomyelins**

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 Nalessso et al. (Nalessso, 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 on 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 Mastrokoli et al. (Mastrokoli, 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).
**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](http://www.absciex.com), Laborjournal 1-2/2010 Geräte für Proteomics, [www.directindustry.com](http://www.directindustry.com)
**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](http://www.absciex.com), Laborjournal 1-2/2010 Geräte für Proteomics, [www.directindustry.com](http://www.directindustry.com).

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


Eichhorst, J. C., M. L. Etter, N. Rousseaux and D. C. Lehotay (2009). "Drugs of
from https://www.cif.iastate.edu/mass-spec/ms-tutorial.


Pusch, W., J. H. Wurmbach, H. Thiele and M. Kostrzewa (2002). "MALDI-TOF mass
spectrometry-based SNP genotyping." Pharmacogenomics 3(4): 537-548.


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


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:


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


Marta Jaremek und Karsten Hiltawsky

Danksagung


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.