CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Publications from Karolinska Institutet

From DEPARTMENT OF ONCOLOGY-PATHOLOGY
SCIENCE FOR LIFE LABORATORY
Karolinska Institutet, Stockholm, Sweden

EXPLORING DISEASE BIOMARKERS AND MECHANISMS USING METABOLOMICS

Anna Lindahl

Stockholm 2017

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by AJ E-print AB
© Anna Lindahl, 2017
ISBN 978-91-7676-611-8

EXPLORING DISEASE BIOMARKERS AND MECHANISMS USING METABOLOMICS

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Anna Lindahl

Principal Supervisor:
Ph.D. Anders Nordström
Umeå Universitet
Department of Molecular Biology

Co-supervisors:
Ph.D. Rainer Heuchel
Karolinska Institutet
Department of CLINTEC

Prof. Janne Lehtiö
Karolinska Institutet
Department of Oncology-Pathology

Ph.D. Roland Nilsson
Karolinska Institutet
Department of Medicine

Opponent:
Ph.D. Elizabeth Want
Imperial College London
Department of Surgery and Cancer

Examination Board:
Prof. Henrik Antti
Umeå Universitet
Department of Chemistry

Ph.D. Svetlana Bajalica Lagercrantz
Karolinska Institutet
Department of Oncology-Pathology

Ph.D. Ali Moazzami
Sveriges Lantbruksuniversitet
Department of Molecular Sciences

Till Jakob och Hannes. Jag är så glad att jag har er!

ABSTRACT

A metabolome is the complete set of small molecules or metabolites present in a biological system. It is the result of internal, genetically determined processes, as well as external factors. The metabolome can consequently be seen as an interface between the genome and the environment. In contrast to gene transcription and translation which do not necessarily result in active gene products, metabolite levels are a direct consequence of the active phenotype and can therefore provide knowledge about cellular mechanisms in health and disease.

The aim of untargeted metabolomics is large-scale detection and quantification of the complete metabolome in a given sample. On a physicochemical level, the metabolome constitutes a complex mixture of compounds, making complete metabolome coverage an analytical challenge. Liquid chromatography-mass spectrometry (LC-MS), which was used throughout this work, is a widely used analytical platform for metabolomics due to its high sensitivity and large metabolome coverage.

In **paper I** of the present thesis, we identified novel potential metabolite markers for pancreatic cancer, comparing serum and plasma samples from patients with pancreatic ductal adenocarcinoma (PDAC) and chronic pancreatitis (CP). The comparison with chronic pancreatitis is not only clinically relevant, since distinguishing CP and PDAC is a challenge with current diagnostic tools; CP also constitutes an inflammatory control condition of the pancreas and therefore aids in the exclusion of non-specific, general disease markers. The comparison with relevant control groups is an important aspect of biomarker discovery study design in general. In **paper II** we showed that the inclusion of non-related disease controls, apart from organ-specific inflammatory controls, can contribute further to the identification of disease-specific biomarkers. We compared the serum metabolic profiles of three non-related diseases with healthy controls and based on overlap analysis of the results we concluded that despite very different etiology and clinical presentation, these three diseases have highly similar effects on the levels of metabolites in serum. In **paper III** we moved from blood-based, systemic metabolic profiles to the intracellular level. Combining LC-MS metabolomics with RNA sequencing, stable isotope tracing and viability assays, we characterized metabolic reprogramming associated with drug-resistance in cancer. When the experimental aim is to discover novel and unexpected dysregulations of the metabolic profile without prior knowledge of the compounds involved, as in **papers I-III**, large metabolome coverage is a key aspect. In **paper IV** we therefore evaluated the impact of the reconstitution solvent on metabolome coverage. Taken together, **papers I-IV** show the large potential of untargeted LC-MS metabolomics as a tool for discovery, to be used as a starting point to trace a chain of molecular events to its origin.

LIST OF SCIENTIFIC PAPERS

I. **Lindahl A**,Heuchel R, Forshed J, Lehtiö J, Löhr M , Nordström A. Discrimination of pancreatic cancer and pancreatitis by LC-MS metabolomics. *Manuscript in revision*

II. **Lindahl A**, Forshed J, Nordström A. Overlap in serum metabolic profiles between non-related diseases: Implications for LC-MS metabolomics biomarker discovery. *Biochemical and Biophysical Research Communications*, 2016, 478(3)

III. Stäubert C, Bhuiyan H, **Lindahl A**, Broom O, Zhu Y, Islam S, Linnarsson S, Lehtiö J, Nordström A. Rewired metabolism in drug-resistant leukemia cells: A metabolic switch hallmarked by reduced dependence on exogenous glutamine. *Journal of Biological Chemistry,* 2015, 290(13)

IV. **Lindahl A**, Sääf S, Lehtiö J, Nordström A. Tuning metabolite coverage in LC-MS metabolomics using the reconstitution solvent composition. *Manuscript in submission*

Publications not included in the present thesis

Mie A, Laursen KH, Åberg KM, Forshed J, **Lindahl A**, Thorup-Kristensen K, Olsson M, Knuthsen P, Larsen EH, Husted S. Discrimination of conventional and organic white cabbage from a long-term field trial study using untargeted LC-MS-based metabolomics. *Analytical and Bioanalytical Chemistry*, 2014, 406(12)

Stäubert C, Krakowsky R, Bhuiyan H, Witek B, **Lindahl A**, Broom O, Nordström A. Increased lanosterol turnover: a metabolic burden for daunorubicin-resistant leukemia cells. *Medical Oncology,* 2016, 33(1)

CONTENTS

LIST OF ABBREVIATIONS

LC-MS	Liquid chromatography-mass spectrometry
NMR	Nuclear magnetic resonance
GC	Gas chromatography
EI	Electron ionization
PCA	Principal components analysis
OPLS-DA	Orthogonal partial least squares-discriminant analysis
NMR	Nuclear magnetic resonance
GC	Gas chromatography
RP	Reversed phase
HILIC	Hydrophilic interaction chromatography
ESI	Electrospray ionization
PDAC	Pancreatic ductal adenocarcinoma
CP	Chronic pancreatitis
Q-ToF	Quadrupole Time-of-Flight
LTQ	Linear trap quadrupole
QQQ	Triple quadrupole
DNR	Daunorubicin
RNAseq	RNA sequencing
PA	Pantothenic acid
FAO	Fatty acid oxidation
TCA	Tricarboxylic acid
Ox phos	Oxidative phosphorylation
LPC	Lysophosphatidylcholine
VIP	Variable importance for the projection

1 BACKGROUND

1.1 THE METABOLOME

The fundamental cellular process referred to as the central dogma of molecular biology starts with the transcription of genes to messenger RNA (mRNA), followed by the translation of mRNA into proteins. The proteins in turn catalyze the chemical reactions of small molecular substrates and products (Figure 1). This thesis will focus on the endpoint of this process: the small molecules, i.e. the metabolites, of the human body and specifically how they are altered in disease.

Metabolites have a wide range of cellular functions. Among others, they are intermediates in energy metabolism, they are building blocks for cellular structures, and they function as signaling molecules activating cellular pathways. In fact, all cellular components consist of metabolites, in the sense that nucleotides are the building blocks of DNA molecules, and amino acids are the primary building blocks of proteins.

The complete set of metabolites of any system under investigation – e.g. tissue, cell type or bodily fluid – is referred to as the metabolome. The composition of a given metabolome depends not only on the sample type, but also on a combination of internal and external factors. Hence the metabolome can be seen as an interface between genetically determined as well as environmental processes, such as diet and lifestyle.

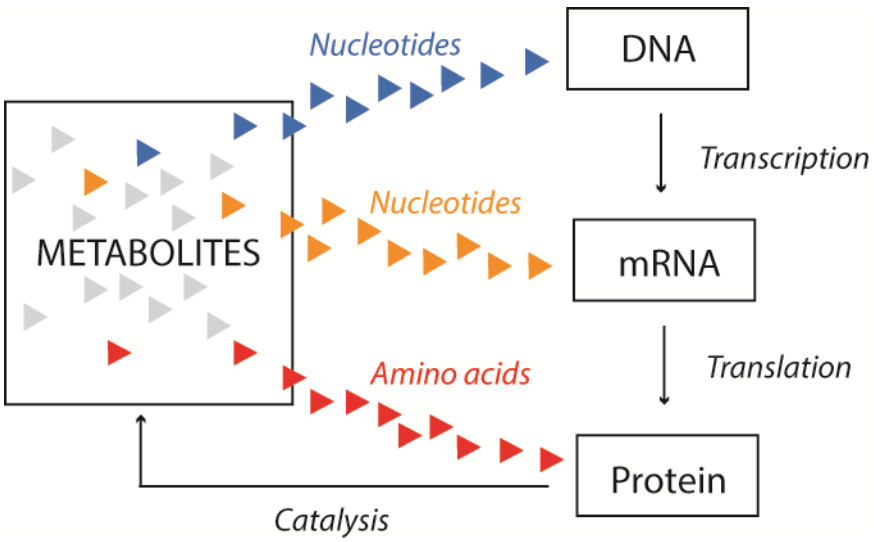

Figure 1. Metabolites and the central dogma of molecular biology. Metabolites can be seen as the endpoint of a cellular process starting at the genome level. At the same time, metabolites are the building blocks of all cellular components. For example, nucleotides and amino acids are the building blocks required for DNA molecules and proteins, respectively.

1.2 METABOLOMICS

1.2.1 Untargeted and targeted quantitative metabolomics

Metabolomics refers to large-scale studies where up to thousands of metabolites can be detected and quantified at the same time in a given sample. The expression "metabolomics" was first introduced in 2000 in the context of plant functional genomics (Fiehn, Kopka et al. 2000).

There are two main approaches to a metabolomics study, targeted and untargeted (Patti, Yanes et al. 2012). In a targeted approach, a hypothesis already exists about the importance of a specific set of metabolites for the research question at hand. In other words the identities or the compound class of the detected metabolites is known beforehand and the analytical workflow can be optimized for that particular set of compounds. The main advantages of this approach are increased analytical depth due to increased sensitivity, higher precision and the possibility to use absolute quantification of metabolite levels.

In untargeted metabolomics on the other hand - the main approach used in this thesis - the identities or the compound classes of the metabolites in a given sample are unknown. Hence the aim is the unbiased detection and quantification of a complete metabolome, usually with the purpose to generate novel hypotheses to be tested at a later stage in the research process. The main advantage of an untargeted approach is the analytical width which allows for the possibility to discover novel, unanticipated dysregulations of metabolite levels. Relative quantification is used to compare the metabolite levels in two or more groups of samples.

Regardless of the approach used, it must be pointed out that both targeted and untargeted metabolomics are quantitative methods that provide limited information about the underlying molecular mechanisms that cause altered metabolite levels. Mechanistic elucidations therefore require follow-up experiments, e.g. using metabolic flux analysis (Johnson, Ivanisevic et al. 2016).

1.2.2 Analytical challenges

As mentioned above, the aim of an untargeted metabolomics study is the unbiased detection and quantification of a complete metabolome. However, this aim is an analytical challenge.

Unlike the proteome, which to a certain extent can be mapped according to the genome, the size of the human metabolome is unknown. Currently, it can be estimated to comprise approximately 6000 metabolites (Wishart, Jewison et al. 2013), of which around 2000 are included in metabolic pathway maps (Thiele, Swainston et al. 2013, Zamboni, Saghatelian et al. 2015). The number of characterized endogenous metabolites is approximately 3000 (Wishart, Jewison et al. 2013), depending on how "endogenous" is defined. Again, unlike the proteome - which is based on combinations of 21 amino acids with similar chemical properties - the metabolome does not consist of a set of common building blocks. Its diversity is instead dictated on the atom level, making it a compound mixture of large complexity as regards physicochemical properties such as hydrophobicity/hydrophilicity (logP) and acid

dissociation constants (pKa). In addition, metabolite concentrations cover a large dynamic range of at least eight orders of magnitude. For example, testosterone can be present in blood at concentrations in the range of 10^{-9}-10^{-12} M, to be compared with 10^{-3} M for glucose (Wishart, Jewison et al. 2013).

1.2.3 Analytical platforms

Due to its complexity, different analytical technologies are needed to detect different parts of the metabolome. Further, metabolome coverage is significantly increased when the detection step is preceded by a metabolite separation method. The two main detectors used in metabolomics are *^{1}H-nuclear magnetic resonance (NMR)* and *mass spectrometry (MS)*. While ^{1}H-NMR requires the presence of hydrogen protons in the molecular structure, MS detects metabolites according to their mass-to-charge ratio (m/z) and therefore requires molecular ionization. Another main difference is that while ^{1}H-NMR analysis is in most cases performed without prior metabolite separation, which limits the number of metabolites detected, MS is often coupled online to a separation system. *Gas chromatography (GC)* and *liquid chromatography (LC)* are the two most common methods of separation used in combination with MS for untargeted metabolomics applications.

Some of the technical and analytical aspects of LC-MS analysis will be discussed more thoroughly below, since this platform was used throughout the present thesis; ^{1}H-NMR and GC-MS however will only be outlined very briefly for the purpose of comparison with LC-MS (Table 1).

Platform	Coverage	Pro	Contra
^{1}H-NMR	- All compounds containing a hydrogen proton	- Structural characterization of unknowns - Reproducible - Absolute quantification - Large dynamic range	- Low sensitivity in complex samples
GC-MS	- Mw < 500 Da - Central carbon metabolism - Sugars/neutrals - Amino acids	- Identification - Small sample amounts needed	- Highly polar compounds excluded - Low Mw only - Identification of unknowns not possible
LC-MS	- All ionizable compounds, i.e. acidic/basic	- High sensitivity - Large coverage - Small sample amounts needed	- High maintenance - More neutral compounds excluded

Table 1. Comparison of major analytical platforms in metabolomics. Mw, molecular weight. Da, Dalton.

1.2.3.1 ^{1}H-NMR

When subjected to electromagnetic radiation of a specific frequency in the presence of a strong magnetic field, each metabolite emits a unique energy signature which depends on the chemical environment of its constituent hydrogen protons. The chemical environment in turn is determined by the position of the hydrogen protons in the molecular structure. Since all endogenous metabolites can be expected to contain at least one hydrogen proton, complete metabolome coverage is in theory possible using ^{1}H-NMR. However, in a complex biological sample containing thousands of compounds – typical for untargeted discovery metabolomics – unique molecular energy signatures become difficult to distinguish, particularly for low-abundant metabolites. ^{1}H-NMR therefore suffers from low sensitivity (Markley, Bruschweiler et al. 2016). Typical compounds detected by ^{1}H-NMR include amino acids, central energy metabolism intermediates and short chain fatty acids (Shi, Brunius et al. 2016). If samples of high purity and concentration are available, ^{1}H-NMR – in comparison with GC- and LC-MS – is the only platform which allows *de novo* structural determination of unknown compounds.

1.2.3.2 GC-MS

In GC, metabolites are separated according to their boiling point. By gradually increasing the temperature a high efficiency separation of a complex mixture is achieved. Molecules with higher molecular weight require higher temperatures to reach their boiling point; higher temperatures are however associated with a risk of thermal metabolite degradation (Fang, Ivanisevic et al. 2015). Similarly, more polar metabolites have higher boiling points, which can be lowered by derivatization of polar functional groups. Derivatization is however performed in a non-aqueous solvent which limits metabolite solubility. In practice, this means that polar compounds as well as compounds with a molecular weight > 500 Da (including derivatization) are rarely detected using GC.

Following GC separation, metabolites are commonly ionized and detected by electron ionization (EI) MS. EI causes extensive metabolite fragmentation in a pattern unique to each compound. The pattern can then be matched against a comprehensive database for metabolite identification. Typical compound classes identified include organic acids, amino acids and more neutral metabolites such as carbohydrates (Dunn, Broadhurst et al. 2011).

1.2.3.3 LC

As previously mentioned, LC can readily be coupled online to MS using atmospheric pressure ionization. LC-MS is arguably the most sensitive and widely used analytical platform for untargeted metabolomics (Buscher, Czernik et al. 2009, Patti, Yanes et al. 2012). It enables the analysis of complex mixtures of large and polar non-volatile compounds in solution, typical for biological samples. Compared to ^{1}H-NMR and GC-MS there are several different LC-MS applications which combined result in the largest metabolome coverage (Nordstrom and Lewensohn 2010, Ivanisevic, Zhu et al. 2013).

The most commonly used LC application in metabolomics is *reversed phase (RP)* chromatography (Yin and Xu 2014), where metabolites are separated according to their hydrophobicity. The liquid sample is injected into a carrier solvent, the mobile phase, which transports the sample through a column, the stationary phase. The stationary phase is hydrophobic, e.g. coated with saturated carbon chains, commonly of 18-carbon atom length (C18); the mobile phase consists of a mixture of H_2O and an organic modifier such as acetonitrile (ACN). Separation is achieved based on analyte partition between the stationary and the mobile phase. Initially, at the time of sample injection, the mobile phase typically contains a high percentage of H_2O. Hydrophobic metabolites will then have a higher affinity for the stationary phase than for the mobile phase and will be more or less retained by the column, causing metabolite separation. Polar metabolites on the other hand are minimally retained, and therefore minimally separated, by the stationary phase. The organic content of the mobile phase is subsequently increased using a linear gradient, causing elution from the column of the more hydrophobic metabolites. Compound classes commonly detected by RPLC-MS are moderately polar/non-polar and non-neutral metabolites such as organic acids (including amino acids), fatty acids and phospholipids.

Hydrophilic interaction liquid chromatography (HILIC) is based on an orthogonal separation mechanism to RPLC. A hydrophilic stationary phase is used, typically in combination with close to 100 % organic solvent as initial mobile phase condition. Polar metabolites are then retained, and therefore separated, to a higher degree than hydrophobic compounds. The aqueous content of the mobile phase can then be gradually increased to elute the more polar metabolites from the column. Typical compound classes detected by HILIC-MS include more polar organic acids, amino acids, nucleotides and intermediates of central energy metabolism (Hemstrom and Irgum 2006).

1.2.3.4 Electrospray ionization

Following LC separation, the solubilized metabolites need to be vaporized and ionized when they are eluted from the column to enable detection by MS. This can be achieved by the application of a high positive or negative voltage (in the range of 2-5 kV) at the MS inlet, which in combination with solvent evaporation leads to the formation of desolvated gas phase ions. This ionization technique is called *electrospray ionization (ESI)* (Cech and Enke 2001).

In contrast to EI, where metabolites in general undergo fragmentation to the point that the molecular ion cannot be detected, the ESI process is a so called “soft” ionization technique which leaves the molecular ion largely intact. The loss or gain of hydrogen protons is the main ionization mechanism in biological samples; due to their small molecular size, metabolites typically have only a single charge site and can therefore lose/gain one proton only, as opposed to e.g. the multiple charge sites of polypeptides. Apart from proton transfer reactions, adduct formation with sodium, potassium and chloride ions is also common.

The ESI response of different metabolites varies depending on their individual properties. Thus two different compounds can have significantly different signal intensity, despite equal

sample concentration. Further, some metabolites ionize in both positive and negative mode ESI – referring to the negative or positive voltage applied at the MS inlet – while others can be detected in one mode only. The highest metabolome coverage is therefore achieved by sample analysis in both positive and negative mode ESI (Nordstrom, Want et al. 2008, Ivanisevic, Zhu et al. 2013).

1.2.3.5 Ion suppression

An analytical concern using LC-MS is ion suppression, commonly as a result of matrix effects. This occurs when the compound(s) of interest co-elute with other – detected or undetected – molecules which alter the ionization efficiency at the ion source. This is more or less inevitable when a complex sample is analyzed since complete compound separation cannot be achieved with any LC method. The effect can be a loss of sensitivity, sometimes to the extent that a compound is undetectable, although sensitivity enhancement is also possible. Further, reproducibility and precision can be affected (Furey, Moriarty et al. 2013).

Matrix effects are to a large extent dependent on sample type. In a typical untargeted metabolomics study design, two or more groups are compared using the same type of sample in both groups. It is therefore reasonable to assume that any matrix effects are fairly equal in both groups and consequently will not lead to false discoveries. The effects are however important to consider for the purpose of increased metabolome coverage, since it is likely that ion suppression reduces overall sensitivity and thus the number of species detected.

1.2.3.6 MS

As previously mentioned, MS is an analytical technique where molecules are separated according to their mass-to-charge ratio (m/z). Essentially, it provides the molecular weight of ions which can be used to determine the elemental composition, structure and identity of a compound.

An MS instrument consists of the three major modules ion source, mass analyzer and detector. The ion source produces gas phase ions, e.g. by ESI as discussed above, while the m/z separation occurs in the mass analyzer. In the final detector module, the number of ions with identical m/z is measured to provide an intensity value. If the MS analysis is preceded by a separation step such as LC, there will also be a retention time value, apart from the intensity, for each m/z. LC-MS data can therefore be described as three-dimensional (Figure 2). Depending on instrument- and analysis type, a fragmentation step can be performed prior to m/z separation and detection. Ions of a particular m/z are isolated and fragmented by collision energy. The m/z values and relative intensities of the fragments can then be used for compound identification.

The performance of a mass analyzer depends to a high degree on its *mass accuracy* and *resolution*. The mass accuracy is calculated by comparing the experimental and theoretical mass value of a known compound, usually expressed in parts per million (ppm). High mass accuracy is important for compound identification and is closely linked to the resolution: if two different compounds of very similar mass value cannot be separated, they will be

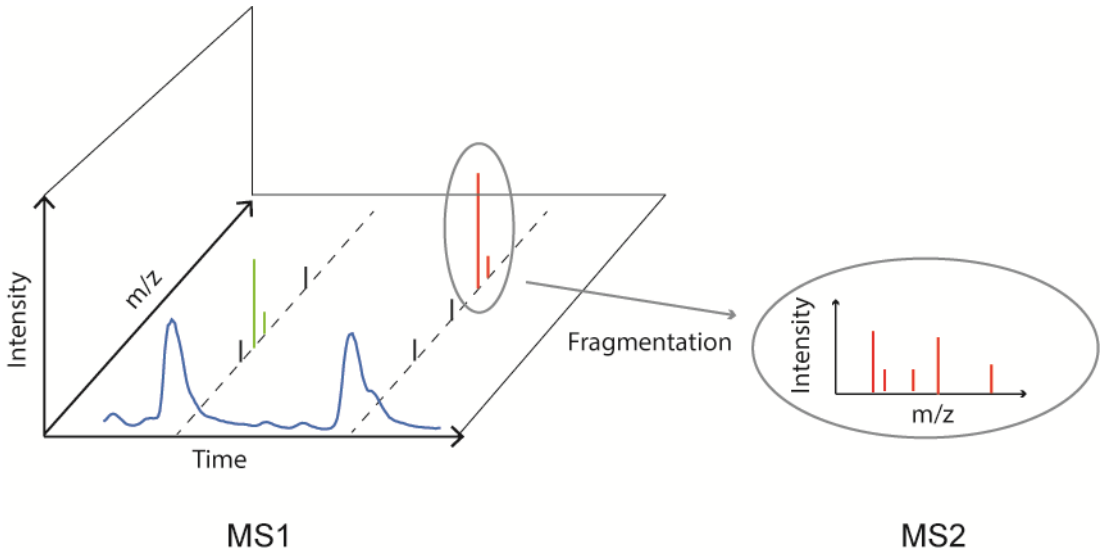

Figure 2. Schematic representation of LC-MS data. The raw data output from an LC-MS analysis can be described as three-dimensional with m/z, retention time and intensity on the axes, also referred to as MS1. An ion of interest (red) can be isolated and fragmented for the purpose of identification, resulting in a fragmentation spectrum referred to as MS2.

combined into one compound only with an average mass value, leading to low mass accuracy.

The two types of mass analyzer used for untargeted metabolomics are *Quadrupole Time-of-Flight (Q-ToF)* and *Linear Trap Quadrupole (LTQ) Orbitrap*. In a Q-ToF, molecular ions are separated according to the time it takes for each ion to pass through a flight tube and reach the detector. The flight time is proportional to the m/z value. The quadrupole that precedes the flight tube can be used for isolation and fragmentation, after which the m/z values of the fragments are determined by the time-of-flight analyzer. Q-ToF-instruments have high linear dynamic range combined with high acquisition rate, i.e. the time required for ion collection, analysis and detection is short; a property which is of importance in LC-MS applications in order to detect a maximal number of compounds as they elute from the LC column.

In an LTQ Orbitrap, ions are separated based on their resonance frequency in an electromagnetic field, which in turn depends on their m/z ratio. In contrast to a Q-ToF, an LTQ Orbitrap can perform multiple rounds of sequential fragmentation of a molecular ion for structural elucidation purposes (see section on metabolite identification below for further information). LTQ Orbitrap analyzers are also capable of very high mass accuracy and resolution.

When the identity of the compound of interest is known, i.e. in targeted applications, a *triple quadrupole (QQQ)* MS is preferable. This type of mass analyzer is based on the stable trajectory of a molecular ion and its fragments through an oscillating electric field. Although resolution and mass accuracy of a QQQ is low compared to the analyzers mentioned above, it has the advantage of high sensitivity and specificity.

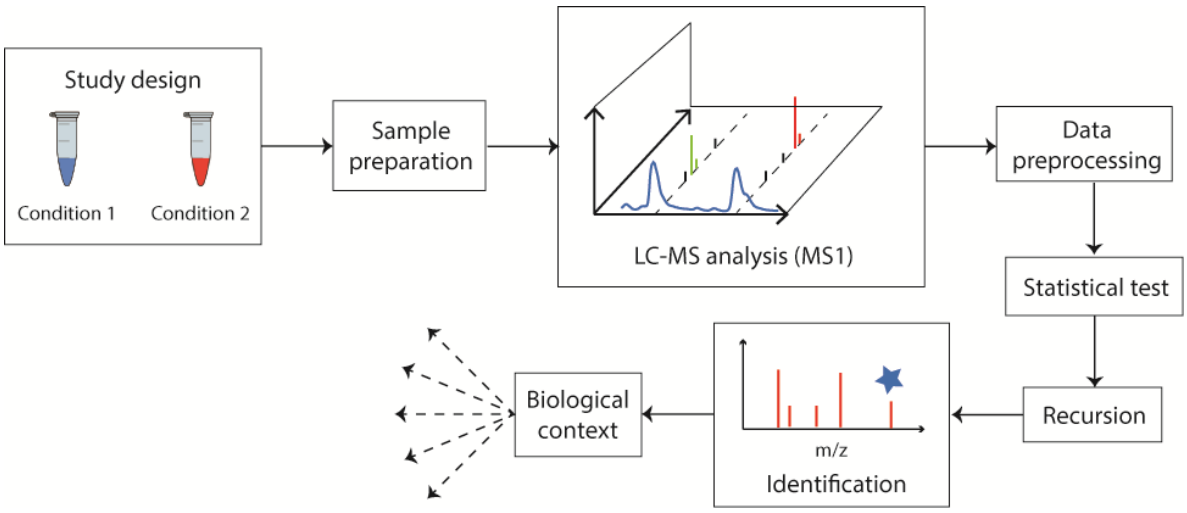

Figure 3. Schematic of a typical LC-MS based untargeted metabolomics workflow. Following study design, sample collection and preparation, LC-MS analysis collecting MS1 data is performed. Data is then preprocessed, including e.g. peak picking and filtering, prior to uni- or multivariate statistical analysis. Reintegration in raw data, also referred to as recursion, in combination with manual curation is an important quality control step. Metabolites are then identified by accurate mass and retention time, combined with MS2 fragmentation spectra for structural elucidation, to be placed in a biological context. Further analyses such as validation or mechanistic investigations can then be performed depending on the research question.

1.2.4 Untargeted metabolomics workflow

A typical LC-MS based untargeted metabolomics workflow aims to identify differences in metabolite levels between different conditions in a particular biological context. The different steps of the untargeted workflow are presented as a schematic in figure 3 (some aspects will also be discussed further in the following two sections).

The first step of the workflow is the study design where one important aspect is to decide upon relevant control(s) for the condition to be studied. Samples are then acquired, prepared by metabolite extraction and screened on the MS1 level to acquire peak integration data for relative quantification. Thousands of peaks are usually detected in each sample.

Following acquisition the MS raw data is processed, including automatic peak picking, alignment and integration. Each peak is annotated with an m/z value and a retention time and is at this stage referred to as a metabolite feature.

Next, a uni- or multivariate statistical test is performed to identify the metabolite features that discriminate between the different conditions studied. Targeted re-integration, i.e. recursion, of the discriminating features is then performed as a feature quality control step. Ideally, recursion is performed prior to the statistical test, but since a certain amount of manual curation is required for quality control, a means of metabolite feature ranking and selection can be necessary already before recursion.

For identification, select or pooled samples are re-analyzed in targeted MS2 mode to acquire fragmentation spectra for structural information. The MS2-spectra and retention time values are ideally matched against a database (in-house or public) together with accurate mass values of the molecular ions. Identified metabolites can then be placed in a biological context e.g. by pathway analysis.

Depending on the research question, the metabolites identified in this initial screening process can then be confirmed as discriminative in a validation sample cohort as part of e.g. a biomarker study; another example is to use the results from the screening process as a hypothesis-generating basis for further mechanistic investigations.

1.2.5 Data analysis

1.2.5.1 Data preprocessing

As mentioned above, a number of data processing steps are needed to extract the relevant information from an MS raw data file in an untargeted metabolomics study. They can be performed automatically using either open-source software like XCMS (Smith, Want et al. 2006) or a vendor-specific version.

Peak picking is the first processing step where the peaks are identified and separated from the background chemical noise. *Peak alignment* due to retention time drift is a processing step that may be necessary to ascertain that the same peak is compared across all samples for each feature. This is particularly important in larger studies where samples are analyzed over long periods of time. *Compound identification* removes the large number of redundant peaks arising mainly from isotopic patterns and adduct formation. *Peak filtering*, where e.g. low abundant metabolite features are excluded, reduces the number of false positives and the amount of manual curation necessary at later stages. *Sample-wise normalization* can be performed to correct for MS signal intensity drift over time. Since relative quantification depends on the correct measurement of differences in metabolite levels that reflect a particular phenotype, normalization is an important aspect of untargeted metabolomics studies. Several different approaches can be used including global and quality control-sample based normalization (Dunn, Broadhurst et al. 2011, Trezzi, Vlassis et al. 2015).

1.2.5.2 Univariate statistical analysis

Group-wise statistical analysis is used to identify variables, i.e. metabolites, that are significantly different in concentration level between two or more sample groups.

The most commonly used univariate statistical analysis method is the Student's t-test, which compares the distribution of one variable at a time between two sample groups. Student's t-test assumes normal distribution and equal variance. If the sample groups have unequal variance, e.g. due to differences in sample size, Welch's unequal variance t-test can be used instead.

A typical untargeted metabolomics experiment produces over a thousand variables to be tested, which leads to an increased probability that some of them differ between the two sample groups by chance alone (i.e. false positives). One commonly used method for controlling the number of false positives is the Benjamini & Hochberg false discovery rate (FDR) correction method. Typically, it results in corrected (higher) p-values, thus decreasing the number of variables passing the significance cut-off level.

1.2.5.3 Multivariate analysis

As opposed to the univariate comparison of one variable at a time, multivariate methods take all variables into account simultaneously. The multivariate approach identifies metabolite patterns, rather than single metabolites, that differ between sample groups. An example of such a metabolite pattern could be compounds that belong to the same dysregulated pathway.

Depending on the research question, there are unsupervised and supervised multivariate methods. Principal components analysis (PCA) is an unsupervised method that identifies the largest variation in the data set regardless of sample groups. It can be used to get an overview of the data, e.g. to assess clustering and outliers. Orthogonal partial least squares-discriminant analysis (OPLS-DA) (Trygg and Wold 2002) on the other hand is a supervised method that identifies the variation that is related to a predefined classification. Comparing two sample groups, OPLS-DA separates the variables that differ between the groups, i.e. the biologically relevant information, from those that vary due to unrelated causes, e.g. the technical analysis.

1.2.6 Metabolite identification

1.2.6.1 Metabolite databases

After selecting the metabolite features that are of interest to the research question, metabolite identities are required to place these features in a biological context. Ideally, high confidence identification is achieved by matching experimental values for *i)* the accurate mass of the molecular ion of interest, *ii)* its retention time and *iii)* MS2 fragmentation spectrum against the corresponding values for the standard metabolite compound. To enable a retention time match, both compounds must be analyzed using the same LC-MS platform. A retention time match in combination with the accurate mass measurement (< 5 ppm) is in most cases specific enough to confirm metabolite identity; an MS2-spectral match confirms the identity even further.

Since the retention time is an important part of the identification process that depends on the analytical platform used, an in-house database containing the retention time for standard metabolite compounds is preferable. There are also a number of public databases available including the Human Metabolome Database (HMDB) (Wishart, Jewison et al. 2013), METLIN (Smith, O'Maille et al. 2005) and MassBank (Horai, Arita et al. 2010). A public database search on accurate mass typically returns a number of possible compound hits that can be used as a starting point for identification. Depending on compound characteristics such as origin and hydrophobicity, some hits can be excluded based on accurate mass alone.

Even if public databases cannot provide retention time information, many compounds are annotated with MS2-spectra. Complete spectral matching, i.e. of both fragment m/z values and intensities, is however difficult to achieve unless the database- and the experimental spectrum have been acquired on the same type of MS instrument.

1.2.6.2 Challenges in metabolite identification

Novel metabolites are still being discovered and the human metabolic pathway map is far from complete. This is apparent in any given untargeted LC-MS metabolomics study, where a large number of peaks remain unidentified and uncharacterized.

Part of the explanation is *in-source fragmentation*. Certain compounds fragment spontaneously during the ionization process in the ESI source. If the individual fragments are ionized, they are detected and can be grouped together based on identical retention time and chromatographic peak profile; typically their MS2-spectra also contain identical fragments. In the case of neutral loss, i.e. when a compound fragment remains unionized, the mass of the molecular ion cannot be reconstructed and identification is almost impossible.

Assuming that the molecular ion is detected, MS2-spectra can be more or less informative depending on the compound class. For example, lysophosphatidylcholines (LPCs) have a typical fragmentation pattern which in combination with retention time and accurate mass results in high confidence identifications. Other compounds are either very hard to fragment, yielding no structural information; while others have a too unspecific fragmentation pattern.

In some cases, additional structural information can be acquired by sequential fragmentation (MS^n) of the MS2 daughter fragments (also referred to as a spectral tree), which in turn may lead to compound characterization (Vaniya and Fiehn 2015).

Provided that the purity and concentration of the compound sample is high enough, NMR is the only analytical platform that allows *de novo* structural identification of an unknown compound, as mentioned above. Such a sample is however very difficult to obtain in the case of a study on clinical patient material.

1.3 METABOLIC REPROGRAMMING IN CANCER

While the transcription of a gene or the translation of mRNA into protein does not necessarily result in an active gene product, the metabolite levels directly reflect the functional phenotype of a cell. Consequently, untargeted metabolomics can be used to discover and to a certain extent investigate novel phenotypes connected to a specific disease, based on the detection of perturbations in metabolite levels.

1.3.1 Cancer development

Cancer is a multifactorial disease where genetic alterations occur over time in a sequential fashion, transforming normal cells into cancer cells. There are six tumor characteristics that have been established as necessary for malignant transformation: 1) sustained proliferative

signaling; 2) evasion of growth suppressors; 3) cell death resistance; 4) replicative immortality; 5) induction of angiogenesis; 6) activation of invasion and metastasis (Hanahan and Weinberg 2000). An additional two characteristics, evasion of immune destruction and reprogramming of energy metabolism, have been suggested and can be considered established as well (Hanahan and Weinberg 2011). The latter characteristic, reprogramming of energy metabolism, is an active field of research where metabolomics analyses are clearly motivated. Below follows a brief outline of the field and a discussion of the role of metabolomics.

1.3.2 Cancer cell metabolism

In order to survive and proliferate, tumor cells undergo metabolic reprogramming. This affects the interactions between cells and metabolites in a number of different cellular compartments. At the cell surface, the metabolite uptake is increased; in the cytosol and mitochondria, metabolites function as building blocks for *de novo* biosynthesis to enable cell proliferation; in the nucleus, metabolites are cofactors or substrates for enzymes involved in epigenetic regulation of gene expression. The metabolic reprogramming of cancer cells can also affect other cell types in the microenvironment (Pavlova and Thompson 2016).

Two metabolites which are important in the biosynthesis required for cellular proliferation are glucose and glutamine (Figure 4). Both show increased uptake in cancer. Regarding glucose, a major advantage of upregulated glycolysis is that it provides a number of carbon intermediates that can be diverted to biosynthetic pathways. For example, glucose-6-phosphate is a starting point for nucleotide synthesis, while 3-phosphoglycerate can be used for the generation of NADPH, which in turn is used as a reducing agent for the generation of e.g. fatty acids and cholesterol.

Regarding glutamine, the exact mechanisms of its role in cancer metabolism are still being investigated. It is however established as an important source of TCA cycle intermediates and nitrogen. Exogenous glutamine is e.g. a key molecule in the synthesis of nucleotides and can also be used in the production of nonessential amino acids. Further, glutamine can provide the starting material for cytosolic acetyl-CoA via the TCA cycle. Acetyl-CoA, just like NADPH, is needed for lipid and cholesterol biosynthesis, molecules which are building blocks for the cellular membranes required for cell proliferation (Altman, Stine et al. 2016).

On the genome level, α-ketoglutarate and acetyl-CoA are examples of metabolites which are important substrates/cosubstrates and cofactors in epigenetic regulation and posttranslational modifications of proteins. Acetyl-CoA provides the acetyl-group for protein acetylation, including histones where acetylation is associated with the activation of gene transcription. α-ketoglutarate is a cosubstrate for a class of enzymes which includes histone and mRNA demethylases (Pavlova and Thompson 2016, Vander Heiden and DeBerardinis 2017).

Metabolites secreted by proliferating cancer cells can change the environment to such a degree that other cell types are affected, including immune cells. Examples are the extracellular accumulation of lactate as a result of upregulated glycolysis which decreases the

immune cell response, and the secretion of kynurenine as a result of increased tryptophan metabolism that can lead to tumor progression as well as immunosuppression (Galluzzi, Kepp et al. 2013, Pavlova and Thompson 2016).

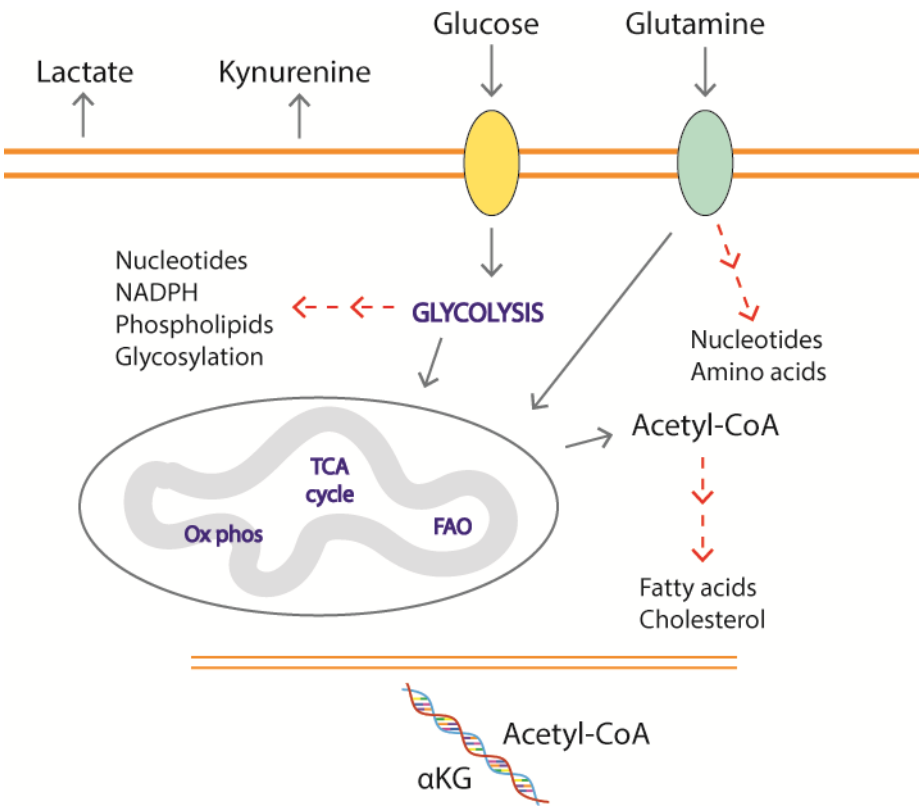

Figure 4. Examples of metabolic reprogramming in cancer cells. Increased uptake of glucose and glutamine provides building blocks for the biosynthesis of e.g. nucleotides, phospholipids and fatty acids required for cellular proliferation, via glycolysis and the TCA cycle, among others. In the nucleus, acetyl-CoA and α-ketoglutarate (αKG) affect cellular transcription levels via epigenetic regulation. Secreted lactate and kynurenine affect the extracellular environment leading to e.g. suppression of the immune system. Ox phos, oxidative phosphorylation; FAO, fatty acid oxidation.

1.4 METABOLOMICS IN BIOMARKER DISCOVERY

Based on the assumption that the perturbations of intra- and extracellular metabolite levels in e.g. tumorigenesis can be detected on the system level, untargeted metabolomics has become widely used in the field of disease biomarker discovery. However, small molecular biomarkers and their detection by MS is obviously not a novel concept (Dalgliesh, Horning et al. 1966, Horning and Horning 1971, Pauling, Robinson et al. 1971); well-known examples routinely used in the clinic include glucose as a marker of diabetes and phenylalanine as a marker of congenital metabolic disorder. The discovery of novel biomarkers, e.g. for early diagnosis or disease prognosis, could improve treatment in several other disease areas as well, including cancer.

1.4.1 Novel metabolite biomarkers

Metabolomics, just as genomics or proteomics, is clearly contributing to the field of biomarker discovery. One example is trimethylamine *N*-oxide which has been identified as a marker of cardiovascular disease (Koeth, Wang et al. 2013). Other examples include a panel of the three amino acids isoleucine, tyrosine and phenylalanine as markers of future development of type 2 diabetes (Wang, Larson et al. 2011); and isoleucine, leucine and valine as markers of increased risk for pancreatic cancer development (Mayers, Wu et al. 2014). Further, the D-enantiomer of 2-hydroxyglutarate as a marker of IDH1-mutated gliomas is currently in clinical trials (Andronesi, Rapalino et al. 2013).

1.4.2 Biomarker discovery study design

Biomarker discovery has proven to be a challenging field of research. Although there have been several successful applications, none of the omics technologies have led to the large advances that were originally hoped for. As a metabolomics case in point, sarcosine was previously reported as a potential biomarker for prostate cancer progression (Sreekumar, Poisson et al. 2009). However, the results could not be validated in further clinical studies using comparable patient samples (Jentzmik, Stephan et al. 2010, Jentzmik, Stephan et al. 2011).

Part of the challenge of metabolomics biomarker discovery is the large inter-individual variation in metabolite levels as a result of genetic and environmental factors (Suhre, Shin et al. 2011, Suhre and Gieger 2012, Dunn, Lin et al. 2015). Contrary to shot-gun proteomics, the analysis of large sample cohorts – in order to balance the variation with more statistical power – is technically possible in metabolomics; the limiting factor is often the access to clinical material. Another important aspect is to obtain detailed patient information, partly to avoid known confounding factors such as smoking and body mass index (BMI), partly to decrease the variation within the sample groups by stratification according to e.g. disease stage.

Apart from acquiring patient samples from the disease of interest, the acquisition of relevant control sample group(s) is an important part of the design of a biomarker study. To avoid the identification of general markers of disease rather than disease-specific markers, the control samples should include an inflammatory control, e.g. pneumonia as a control for lung cancer (Chechlinska, Kowalewska et al. 2010, Kowalewska, Nowak et al. 2010). The need to control for non-specific inflammatory mechanisms is an established fact in proteomics (Petrak, Ivanek et al. 2008) but is not always implemented in metabolomics.

Traditionally, potential biomarkers have been identified by univariate statistical analysis as single, independent discriminative entities. However, considering that metabolites are highly likely to be part of a series of enzymatic reactions dependent on each other, the use of multivariate analysis methods has become increasingly common in biomarker discovery studies; it has also been shown that a panel of biomarkers can increase biomarker specificity (Wang, Larson et al. 2011).

2 THE PRESENT THESIS

2.1 AIMS

The overall aim of this thesis work was to use untargeted LC-MS metabolomics as a tool for discovery of potential biomarkers and novel disease mechanisms. This included aspects of biomarker discovery study design and analytical method development, towards the goal of acquiring biologically and clinically relevant information.

The specific aims of papers I-IV were:

Paper I: To identify metabolites discriminating between PDAC and chronic pancreatitis in human serum and plasma by untargeted LC-MS metabolomics, for potential clinical use as diagnostic markers.

Paper II: To determine the extent of overlap between the serum metabolite profiles of three diseases in comparison with healthy controls, providing implications for the design of future biomarker discovery studies.

Paper III: To identify transcripts and metabolites differing between drug-resistant and sensitive leukemia cells and investigate the underlying molecular mechanisms.

Paper IV: To investigate the impact of the reconstitution solvent composition on the metabolome coverage in untargeted LC-MS metabolomics.

2.2 MATERIALS AND METHODS

In this section, key methods and analytical aspects of papers I-IV are described. Detailed, complete descriptions of the materials and methods used can be found in each paper.

2.2.1 Samples and study design

Paper I was based on two different sample cohorts. Cohort one, the discovery cohort, consisted of clinical serum samples from pancreatic ductal adenocarcinoma (PDAC, n = 44) and chronic pancreatitis (CP, n = 23) patients collected in Germany. Cohort two, the validation cohort, consisted of plasma samples from PDAC (n = 20) and CP (n = 31) patients collected in Sweden. Starting with the discovery cohort, PDAC and CP samples were compared to identify metabolites differing between the two groups. The results from this comparison were then confirmed in the validation cohort.

Paper II was based on commercially acquired serum samples from patients with the three non-related diseases community acquired pneumonia (n = 25), congestive heart failure (n = 40) and Non-Hodgkin's lymphoma (n = 40), as well as healthy controls (n = 40). Each disease was compared to the healthy control group followed by overlap analysis of the results (Figure 5).

Paper III was based on chemotherapeutic drug-sensitive and drug-resistant human leukemia cell lines. As a first step, the sensitive and resistant cell lines were first compared to identify perturbations on the translational and metabolic level. The results from this initial screening were then used for further investigations of the molecular characteristics of drug-resistance.

Paper IV was based on commercially available complex cell medium samples and commercially available human serum samples from leukemia patients (n = 20) and healthy controls (n = 20). The cell medium samples were used in the initial method evaluation step, followed by method application on the human serum samples comparing patient and control samples.

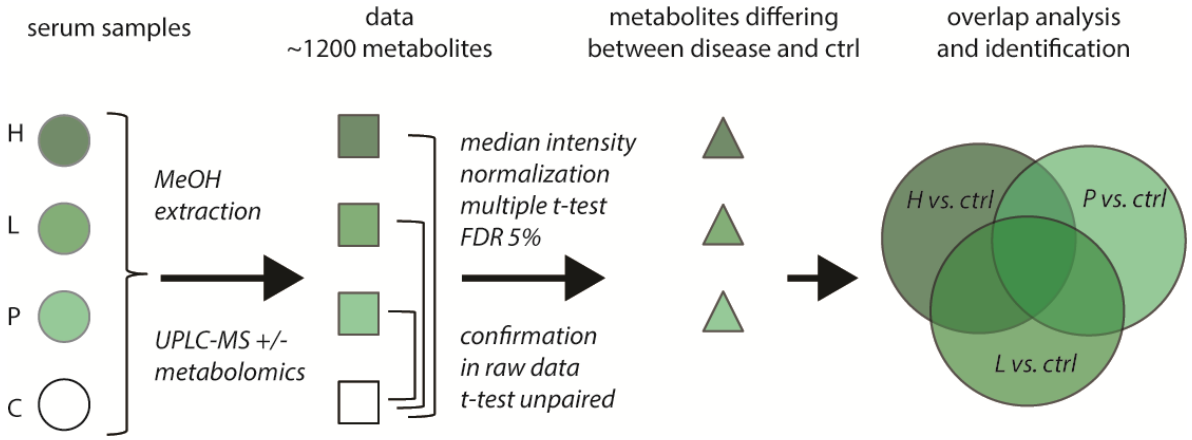

Figure 5. Sample groups and study design in paper II. The three non-related diseases congestive heart failure (H), Non-Hodgkin's lymphoma (L) and community acquired pneumonia (P) were compared individually to a healthy control group. The resulting differential metabolites were analyzed for overlap.

2.2.2 LC-MS instrumentation

In **paper I**, two different LC-MS platforms were used for the analysis of the discovery and the validation cohort. RPLC-LTQ Orbitrap was used for the screening and metabolite identification of the discovery cohort. The validation cohort was analyzed with an RPLC-Q-ToF instrument in untargeted and targeted mode for relative quantification and identification, respectively. The RPLC-Q-ToF platform was also used for untargeted screening in papers **II, III** and **IV**. Additional structural elucidation by pre-fractionation and direct injection MS^n was performed on an LTQ Orbitrap for **paper II**. In **paper III**, a QQQ instrument was used for targeted relative quantification of stable isotope-labelled pantothenic acid, CoA and acetyl-CoA.

2.2.3 Data preprocessing

Peak picking, alignment and integration of the RPLC-LTQ Orbitrap data set from the discovery cohort in **paper I** was performed with the open-source software XCMS. All other data sets were processed using vendor-specific software. Sample-wise median normalization was applied in **papers I, II** and **IV**. For the multivariate data analysis in **papers I** and **IV**, the software default UV-scaling was used in combination with log transformation.

2.2.4 Statistical analysis

Univariate Student's t-test or Welch's unequal variance t-test were performed to identify differing metabolite features in all papers.

In **paper I**, supervised multivariate OPLS-DA was performed using the software SIMCA (Umetrics), in parallel with the univariate approach. The stability and generalizability of OPLS-DA models is evaluated by cross-validation, where a subset of randomly chosen samples is left out. The OPLS-DA model based on the remaining samples is then used to predict which sample group, or class, the samples in the left-out subset belong to. The process is then repeated until all samples have been predicted once.

The selection of discriminant features from the OPLS-DA was made based on the variable importance for the projection (VIP). Features with a VIP-value ≥ 1 and significant within a 95 % confidence interval based on the cross-validation were selected, according to software recommendations. In the initial OPLS-DA of the discovery cohort including 4578 features, the VIP-value cut-off ≥ 1.5 was used in order to limit the number of features selected for manual curation.

Apart from the VIP-value, the cross-validation is also the basis for three other measures of model performance; R2X(cum), Q2(cum) and CV-ANOVA. R2X(cum) is a measure of how much the data set is used to build the OPLS-DA model. Q2(cum) is a measure of how well the model predicts sample class. CV-ANOVA is a measure of the probability that the model is the result of chance only.

In **paper IV**, unsupervised PCA, also performed in SIMCA, was used to detect trends related to the composition of the sample reconstitution solvent.

2.2.5 Raw data curation

Metabolite feature quality control by manual curation, i.e. raw data inspection of extracted ion chromatograms, was an important part of the workflow, above all in **papers I** and **II**. Automatic peak picking, alignment, integration and compound identification is a necessary part of the untargeted metabolomics workflow, but the results need to be verified in the raw data to remove e.g. misaligned and low intensity peaks.

2.2.6 Metabolite identification

Metabolite identities were determined by in-house and METLIN database matching against standard metabolite compounds based on accurate mass, retention time and MS2 spectra in all papers. In **paper II** MS^n fragmentation was also performed to gain further structural information.

2.3 RESULTS AND DISCUSSION

2.3.1 Paper I

In this study, two independent sample cohorts were analyzed by untargeted RPLC-MS metabolomics to identify single as well as a panel of metabolites discriminating PDAC and CP in serum and plasma. Uni- and multivariate analyses were performed in parallel, starting with the discovery cohort data set containing 4578 metabolite features including isotopes and adducts.

In the univariate analysis workflow, features were ranked by Welch's t-test and the top 254 ($p < 0.05$) were selected for further evaluation. FDR correction was not applied at this stage; the Welch's t-test was used only as a tool for metabolite feature ranking and selection. *i)* Acceptable raw data quality, *ii)* presence in the validation cohort and *iii)* known identity were used as a further set of selection criteria, leaving 17 metabolites. 11 of these were confirmed as significant (Welch's t-test, $p < 0.05$) in the independent validation cohort. However, fold-change calculations revealed that eight metabolites, all phospholipids, were regulated in opposite directions in the two cohorts. After exclusion of the phospholipids the three metabolites glycocholic acid, hexanoylcarnitine and N-palmitoyl glutamic acid remained as single discriminative markers for PDAC compared to CP (Figure 6).

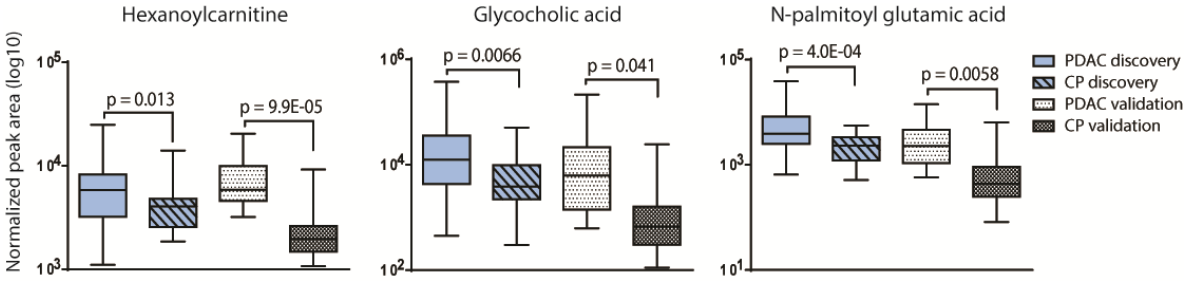

Figure 6. Metabolites discriminating PDAC and CP in serum and plasma, as identified by univariate analysis and validated in a second independent cohort. Statistical test: Welch's unequal variances t-test. Box plot settings: Line at median; range, minimum to maximum.

In the multivariate analysis workflow, OPLS-DA of the discovery cohort data set was used to select 259 discriminative metabolite features, based on VIP-values, for further evaluation. The three exclusion criteria used in the univariate analysis workflow were applied here as well, leaving 19 metabolites. All 19 were confirmed as discriminative in OPLS-DA of the validation cohort. However, as in the univariate analysis, 14 phospholipids were regulated in opposite directions in the two cohorts and were therefore excluded. The five remaining metabolites N-palmitoyl glutamic acid, glycocholic acid, hexanoylcarnitine, chenodeoxyglycocholate and phenylacetylglutamine (PAGN) were used to build an OPLS-DA model to evaluate their performance as a marker panel (Figure 7). As there was a three-metabolite overlap between the results from the multivariate analysis and the univariate, the two approaches corroborated each other.

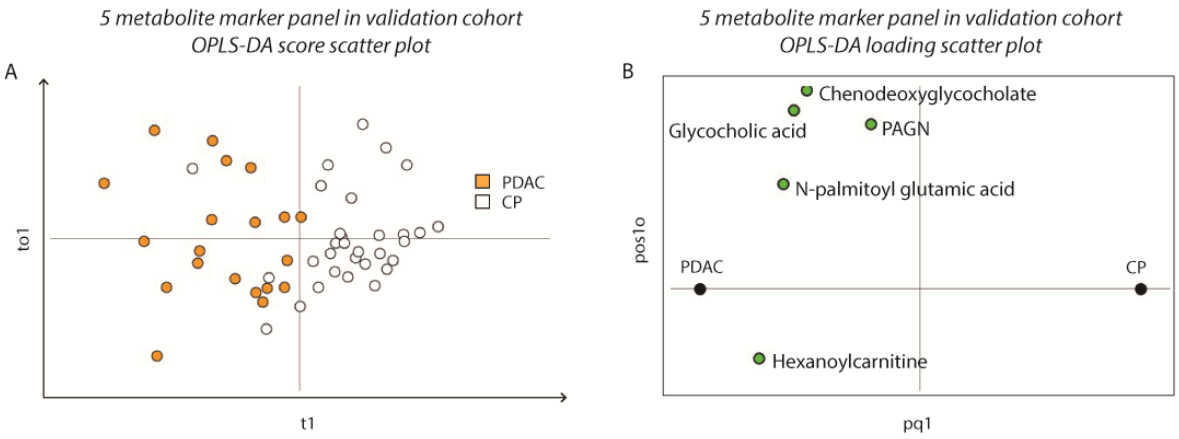

Figure 7. OPLS-DA model of a five-metabolite marker panel in the validation cohort. **A.** Score scatter plot of PDAC (n = 20) and CP (n = 31) samples. The Q2(cum) value was 0.513; i.e. approximately 50 % of samples were correctly classified. The R2X(cum) value was 0.736. Cross-validated ANOVA was p = 8.2E-07. **B.** Loading scatter plot with dummy PDAC and CP variables. The five metabolites are situated to the left of origo, indicating that they are up-regulated in PDAC compared to CP.

Apart from constituting an inflammatory control of the pancreas for PDAC-specific biomarker discovery, CP is a known risk factor for PDAC development. The two conditions are however difficult to distinguish clinically since they present with similar inflammatory symptoms. Screening of CP patients could therefore be motivated as they constitute a limited, well-defined patient group, where a diagnostic biomarker could aid in the early detection of PDAC.

An independent validation cohort, as used in the present work, is an important aspect of a biomarker study as it provides a measure of the robustness and generalizability of the potential markers. Until very recently, there had been no previous untargeted studies comparable to the present one in terms of patient number and validation of results in a second cohort. However, Mayerle *et al*. published a large study (n = 914) comparing PDAC with CP as well as healthy and non-pancreatic disease controls in serum and plasma using a comprehensive analytical platform combining GC- and LC-MS including lipid analysis (Mayerle, Kalthoff et al. 2017). A nine-metabolite marker panel, consisting of lipids and organic acids, was identified and validated in an independent test set; a majority of these metabolites could not have been detected by the analytical platform used in the present study, explaining the lack of overlap in metabolite identities.

The down-regulation of phospholipids in the discovery cohort as opposed to up-regulation in the validation cohort in the present study could possibly be explained by the difference in sample matrices, i.e. serum and plasma, of the two cohorts. The different sample handling methods of serum and plasma have been shown to affect the concentration levels of some metabolites including phospholipids (Yu, Kastenmuller et al. 2011). Ideally, the same sample matrix should have been used in both cohorts. Further, the present study would have been strengthened by detailed patient information on known confounders, e.g. smoking status, and disease stage. Larger clinical studies are needed to evaluate the use of these metabolite markers; still, we believe that they have potential as diagnostic markers for PDAC in CP patients.

2.3.2 Paper II

In this study, we compared Non-Hodgkin's lymphoma, community acquired pneumonia and congestive heart failure with healthy controls using untargeted RPLC-MS metabolomics in positive and negative mode. Results show that the metabolic serum profiles of the three diseases display very similar differences when compared to healthy controls, despite very different etiology and clinical presentation of the three conditions.

Each disease was compared individually to the healthy control group using Student's t-test with Benjamini & Hochberg FDR correction (5 %). After exclusion of features overlapping between ionization modes and manual curation in raw data to exclude low quality metabolite features, a total of 178 differential features remained, of which 66 % were identified.

A majority of the differential metabolites, 61 %, overlapped between two or all three diseases compared to healthy controls (Figure 8). The largest total number of differential metabolites (n = 134) was identified in pneumonia, which also had the largest number of unique metabolites (n = 51). The number of metabolites unique to heart failure and lymphoma were only 12 and 6, respectively.

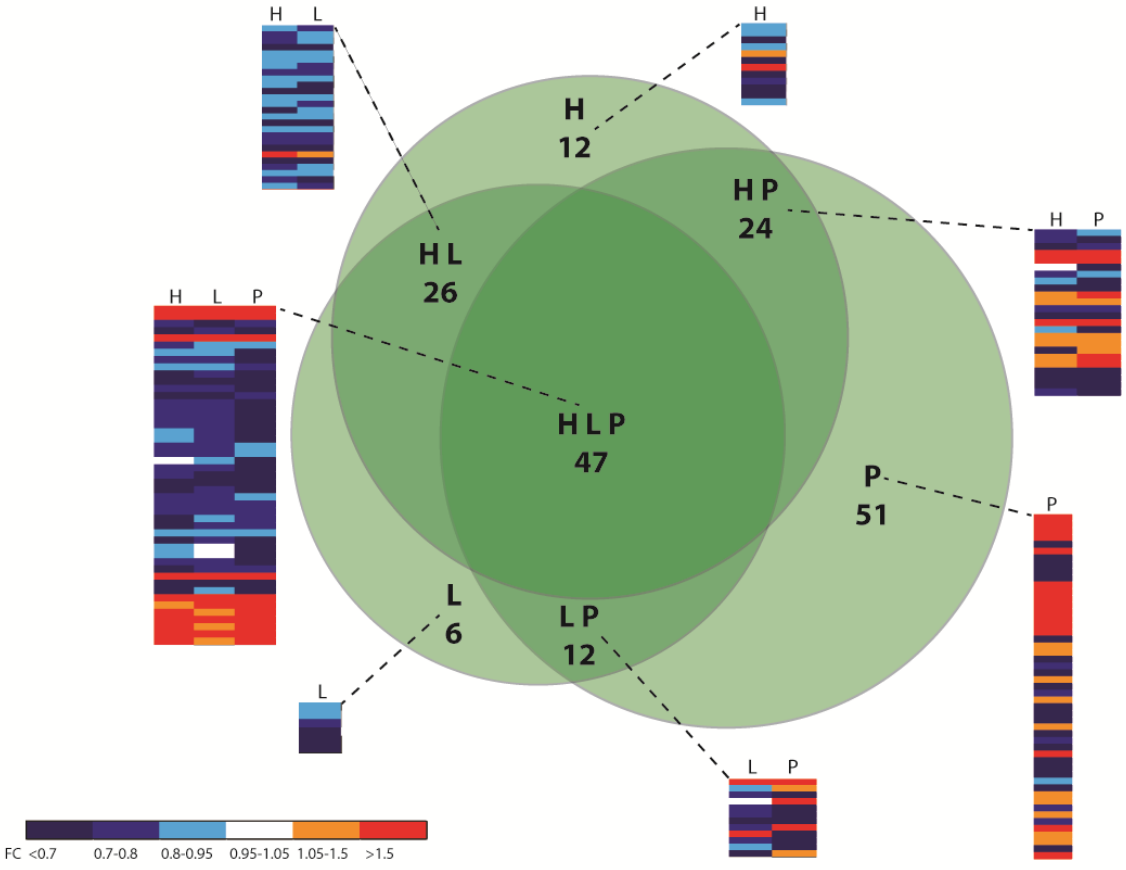

Figure 8. Overlap analysis of the metabolic profiles of the three non-related diseases congestive heart failure (H), Non-Hodgkin's lymphoma (L) and community acquired pneumonia (P) compared to healthy controls. Metabolites differing between disease and control were identified by multiple t-test with Benjamini & Hochberg FDR correction (5 %). 61 % of a total of 178 metabolites were shared between two or all three diseases (HLP, HL, HP, LP in the Venn diagram).Only 12 and 6 metabolites were unique to heart failure and lymphoma, respectively. A majority of the shared metabolites were down-regulated in disease as determined by fold-change (FC) calculations.

The differential metabolites overlapping between all three diseases were to a large extent (43 %) LPCs, which show a strong association to various disease states according to the literature (references are available in paper II, Supplementary Table 3).

We further used our data set for three hypothetical lymphoma biomarker studies using different types of control groups (Figure 9). First, lymphoma was compared to healthy controls, identifying LPC(18:3) as a differential metabolite. Second, different disease stages of lymphoma (stage I-IV) were compared, revealing a significant decrease in the level of LPC(18:3) during disease progression. In the third study design, heart failure and lymphoma were included as non-related disease controls. Using these control groups, it is clear that LPC(18:3) is not a lymphoma-specific metabolite marker.

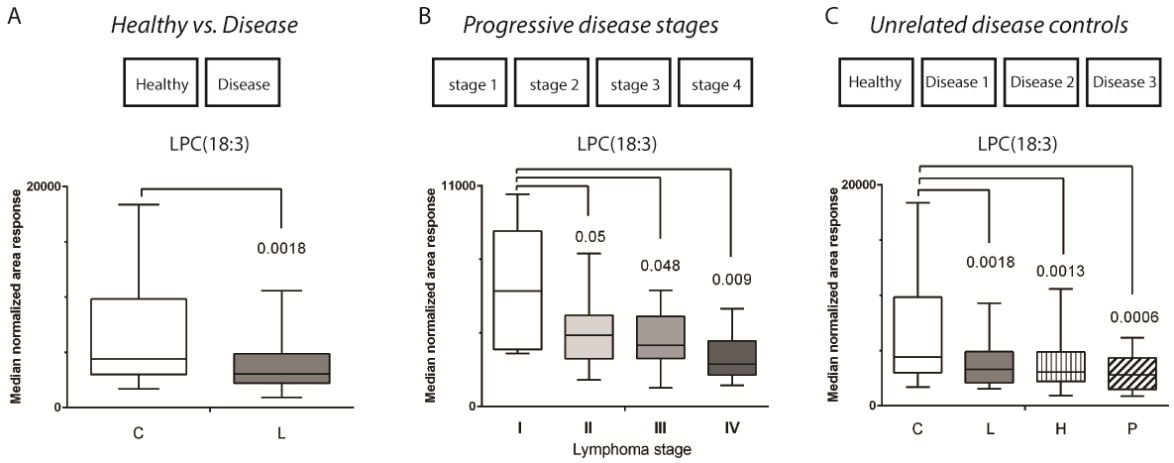

Figure 9. Hypothetical lymphoma biomarker discovery studies using different types of control groups. **A.** When lymphoma (L) is compared to healthy controls (C) only, LPC(18:3) appears to be disease specific. **B.** Comparing different stages of lymphoma, LPC(18:3) levels decrease with disease progression. **C.** When the non-related disease controls congestive heart failure (H) and pneumonia (P) are included, LPC(18:3) is no longer specific to lymphoma. Statistical test: Student's t-test. Box plot settings: range, min. to max.; line at median.

The disease specificity of a potential marker is a key aspect of a biomarker discovery study. Although admittedly the number of unique metabolites identified in this study could have been higher if a second analytical platform such as HILIC-MS had been used, the number of metabolites unique to a particular condition was surprisingly low. In addition, some of these supposedly unique markers have been previously identified as specific to other, non-related conditions as well. A case in point is phenylalanine, unique to lymphoma in the present study, but previously identified as unique to e.g. Alzheimer's disease (Gonzalez-Dominguez, Garcia-Barrera et al. 2015).

LPCs in particular appear unlikely as disease-specific markers, based on this study. Given the fact that LPCs are inflammatory mediators (Sevastou, Kaffe et al. 2013) and inflammation often is part of a general host response in disease (Chechlinska, Kowalewska et al. 2010), it is perhaps not surprising that they have been found in altered levels in a range of different conditions. As shown above (Figure 9), the inclusion of non-related disease controls may contribute to the exclusion of general disease markers such as LPC(18:3). We therefore suggest that future biomarker discovery studies should include control groups from clearly separate disease categories, to increase the chances of successful further validation.

2.3.3 Paper III

In this study, we compared drug-resistant with drug-sensitive human cell lines using RNA sequencing (RNAseq) and untargeted RPLC-MS metabolomics in positive and negative mode. Based on the results from these large-scale screening methods, we investigated underlying molecular mechanisms further using e.g. stable isotope tracing and viability assays to characterize metabolic reprogramming in drug-resistance.

We generated a drug-resistant cell line, CEM/R2, from the acute lymphoblastic leukemia drug-sensitive cell line CEM, by treatment with increasing concentrations of daunorubicin (DNR).

The transcript levels of the resistant and sensitive cell lines were analyzed using RNAseq and compared based on negative binomial distribution (Anders and Huber 2010), identifying 173 differentially expressed mRNA transcripts. 23 of these transcripts were linked to cell metabolism, including cholesterol and glutamine biosynthesis. The changes in cholesterol synthesis were investigated further in a subsequent study (Staubert, Krakowsky et al. 2016) not included in the present thesis.

Probing glutamine metabolism, results showed that the resistant cells not only lacked glutamine synthase activity, but were also resistant to deprivation of exogenous glutamine; on the other hand, they were more dependent on glucose than the sensitive cells. High exogenous glutamine dependence is a common characteristic of cancer cells (Altman, Stine et al. 2016), although the reverse has also been observed (Pavlova and Thompson 2016); the higher glucose dependence observed here may indicate a drug-resistance-linked metabolic shift towards glycolysis as the major provider of building blocks for cellular proliferation.

Next, untargeted metabolomics analysis identified a number of metabolites differing (Student's *t*-test, FDR 5 %) between the sensitive and resistant cell lines (Figure 10). Large fold-change values were observed for e.g. TCA-cycle intermediates, metabolites involved in fatty acid oxidation (FAO) and pantothenic acid (PA).

The level of PA, an essential nutrient and precursor of CoA, was approximately four-fold lower in resistant cells than in sensitive. To investigate whether the decrease in PA levels was caused by diminished uptake or increased intracellular turnover into CoA, sensitive and resistant cells were administered stable isotope-labelled PA. After 24 h, the extracellular level of labelled PA was significantly higher in the resistant cells and the rate of *de novo* CoA synthesis was significantly lower (Figure 11). These results suggest that the resistant cells have a lower capacity for CoA synthesis due to limitations in PA uptake. Considering the central role of CoA in fatty acid metabolism, this was a potential weakness which was tested by treatment with fenofibrate, a blood cholesterol lowering agent which simultaneously increases the rates of FAO and fatty acid synthesis. Results showed that fenofibrate treatment had a larger impact on the proliferation of resistant cells than sensitive, and resistant cells were re-sensitized to DNR when co-treated with fenofibrate.

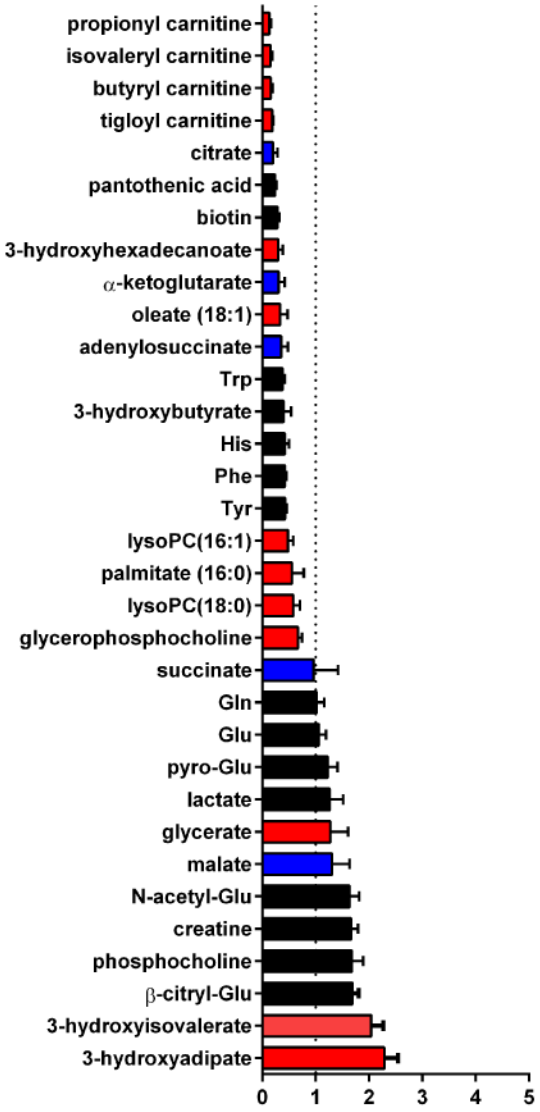

Figure 10. Fold-change of significantly different metabolites (with the exception of succinate, glutamine and glutamate) comparing resistant over sensitive cells, as analyzed by untargeted LC-MS metabolomics. Statistical test: multiple t-test with Benjamini & Hochberg FDR correction (5 %). Metabolites associated with fatty acid metabolism are highlighted in red; TCA cycle intermediates are highlighted in blue.

In this study, large-scale screening using RNAseq and untargeted metabolomics were used as a starting point to characterize the phenotype of a drug-resistant cancer cell line. Metabolic reprogramming involving reduced dependence on glutamine, increased dependence on glucose and decreased fatty acid metabolism was revealed, which could potentially be exploited for the identification of novel drug targets.

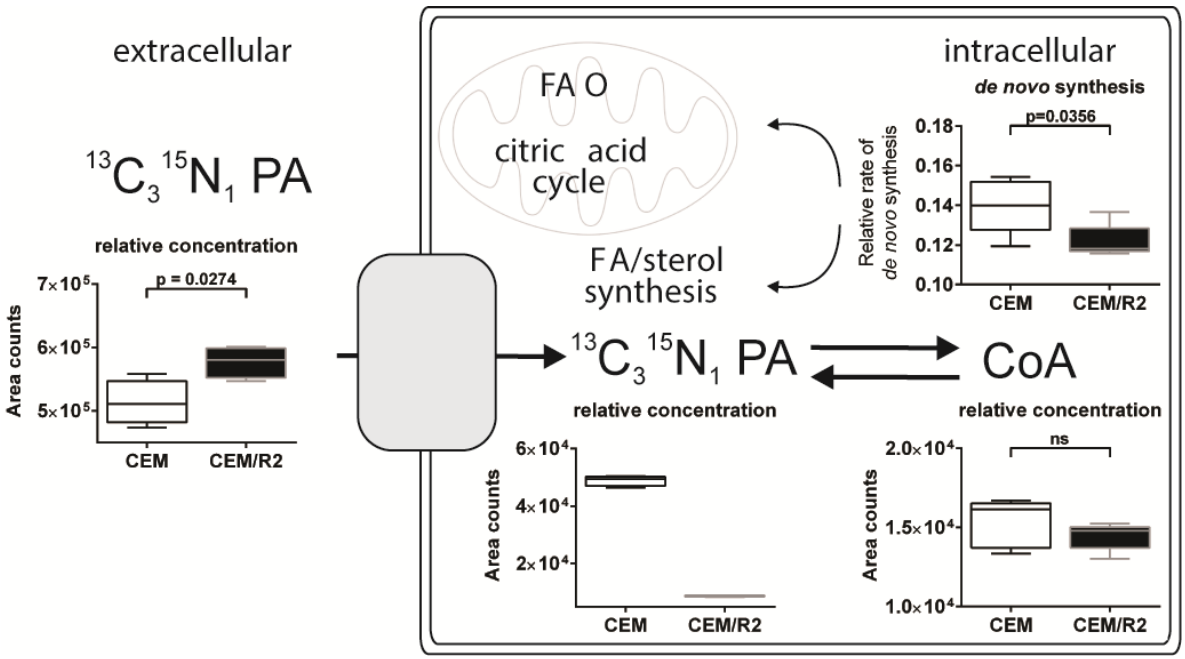

Figure 11. Stable isotope tracing reveals decreased uptake of pantothenic acid (PA), a precursor for CoA, and decreased *de novo* synthesis of CoA in resistant cells. The extracellular levels of stable isotope-labelled PA were higher in resistant cells compared to sensitive, indicating decreased uptake of PA in resistant cells. Intracellular levels of CoA were equal between cell lines but the rate of *de novo* synthesis of CoA was significantly lower in resistant cells. Statistical test: Student's t-test. Box plot settings: range, min. to max.; line at median. FAO, fatty acid oxidation.

2.3.4 Paper IV

In this study, we evaluated the impact of the reconstitution solvent on metabolome coverage in untargeted RPLC-MS metabolomics on aqueous MeOH-extracted biological samples. Our results show that the choice of reconstitution solvent can significantly increase the number of metabolite features detected as well as the metabolome coverage, which in turn may increase the number of significant discoveries in typical biomarker discovery studies.

The effect of the reconstitution solvent was first evaluated on MeOH-extracted complex cell medium, which represents a commercially available, reproducible biological matrix with a complexity similar to that of human serum or plasma. Samples were reconstituted in MeOH:H_2O ratios ranging from 0-100 % MeOH in 10 % increments. 50:50 MeOH:H_2O is a commonly used reconstitution solvent composition in untargeted RPLC-MS (Ivanisevic, Zhu et al. 2013, Want, Masson et al. 2013, Drogan, Dunn et al. 2015, Fang, Ivanisevic et al. 2015), representing an effort to increase the metabolome coverage by including not only hydrophilic, but also more hydrophobic metabolites which are poorly solubilized in pure H_2O.

Using untargeted RPLC-MS analysis the highest number of metabolite features (n = 1491) was detected in samples reconstituted in 100 % H_2O, and then decreasing linearly by each 10 % increment of MeOH (Figure 12). 100 % MeOH resulted in the lowest number of features detected (n = 1099). Lower feature peak heights as a result of decreased solubility and poorer peak shapes as a result of increased solvent strength appeared to be the major reasons for the inverse correlation between MeOH ratio and number of features detected.

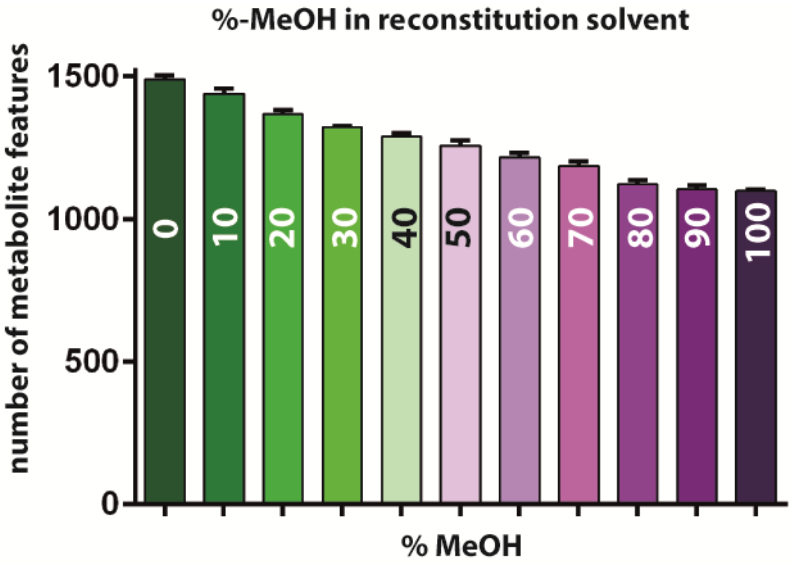

Figure 12. Untargeted LC-MS analysis of samples reconstituted in MeOH:H_2O ratios ranging from 0-100 % MeOH in 10 % increments. Reconstitution in 100 % H_2O results in the highest number of features detected. The differences in number of features between different increments are statistically significant (Student's t-test, $p < 0.05$, 5 replicates) except between 80-90 and 90-100 % MeOH.

Since the polarity of a metabolite affects its solubility, we investigated whether the effect of the reconstitution solvent on metabolite response could be predicted by metabolite polarity, as determined by logP. Standard metabolite compounds representing a wide range of endogenous metabolite polarity (logP -3.3 – 7.1) were spiked in complex cell medium and analyzed by RPLC-MS. Almost all metabolites with logP < 5 increased in response at lower ratios of MeOH, i.e. H_2O was the preferable reconstitution solvent for this group of metabolites. Since 74 % of the metabolites of the known human blood metabolome have a logP < 5 (Figure 13) based on the HMDB (Wishart, Jewison et al. 2013), H_2O is likely to be the preferable reconstitution solvent for a majority of blood metabolites.

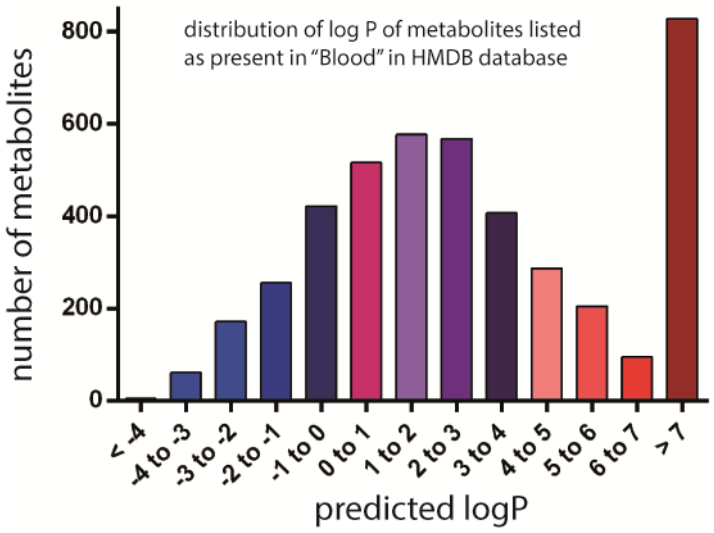

Figure 13. The predicted logP distribution of the known human blood metabolome as extracted from the HMDB. A majority of metabolites (74 %) have a logP < 5, indicating that a highly aqueous reconstitution solvent would be preferable.

Maximizing the metabolome coverage is a central aspect of any untargeted metabolomics study since no prior knowledge about the compound(s) of interest is available. Our results so far suggested that for MeOH-extracted biological samples, 100 % H_2O as reconstitution solvent will maximize metabolome coverage in terms of the number of features detected. In terms of the composition of the detected part of the metabolome, further investigation was needed, especially in comparison with the commonly used 50:50 MeOH:H_2O ratio. The feature lists from analysis of pure MeOH and H_2O were used separately for targeted feature extraction in the 50 % MeOH samples to avoid bias toward either solvent (Figure 14A). Results showed that there appears to be a large overlap in metabolome coverage between 100 % H_2O and MeOH. The addition of 50 % MeOH to pure water resulted in a sensitivity loss for the major fraction of features (57 %), while for the second largest fraction (38 %) the response remained unchanged (Figure 14B). Conversely, the addition of 50 % H_2O to pure MeOH did not affect the response of the majority of features (61 %); in fact, sensitivity increased for the second largest fraction of features (28 %, Figure 14C). The magnitudes of the changes in response were similar in both cases. This suggests that the loss of a more hydrophobic part of the metabolome when reconstituting samples in 100 % H_2O is clearly outweighed by the increased coverage of more hydrophilic features. Further, considering that MeOH-extracted samples constitute a relatively hydrophilic environment already prior to the

reconstitution step, a complementary extraction method e.g. using MeOH/chloroform (Mehlem, Palombo et al. 2016) would be more useful than MeOH in terms of increased hydrophobic coverage.

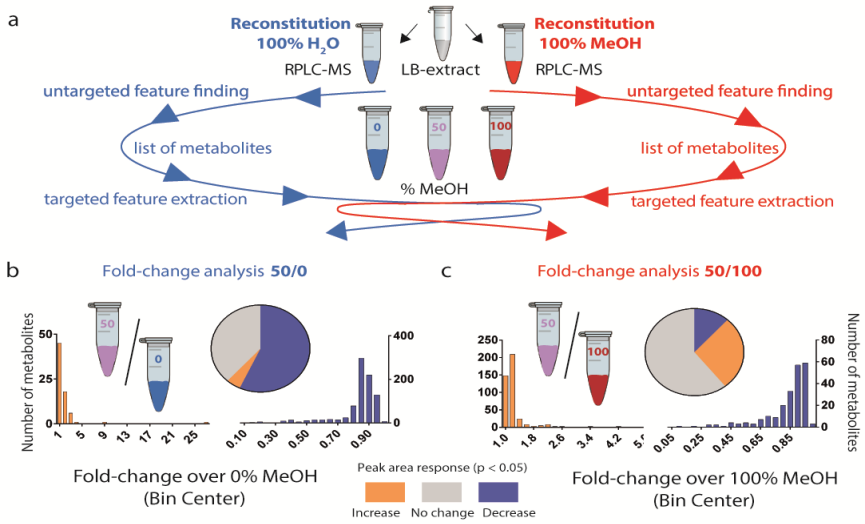

Figure 14. Effect of the reconstitution solvent on the metabolome coverage in terms of the composition of the detected part of the metabolome. **A.** To avoid bias toward either of the pure solvents, the feature lists from samples reconstituted in 100 % MeOH and H_2O were used separately for targeted feature extraction in samples reconstituted in 50 % MeOH. **B.** The addition of 50 % MeOH to pure H_2O results in decreased peak area response for 57 % of metabolite features. The response is unchanged for 38 % of features. **C.** The addition of 50 % H_2O to pure MeOH results in increased response for 28 % of features while 61 % remain unaffected. Statistical test: Student's t-test.

To evaluate the impact of our findings we proceeded with untargeted RPLC-MS analysis of MeOH-extracted human serum samples from lymphoma patients and healthy controls, simulating a basic biomarker discovery scenario. A comparison of the two sample groups reconstituted in 0, 50 or 100 % MeOH by Student's t-test (FDR correction 5 %) resulted in the identification of six significantly differential metabolites when using 100 % H_2O, as opposed to no significant discoveries using 50 or 100 % MeOH.

In conclusion, this study shows that reconstitution of aqueous MeOH-extracted samples in 100 % H_2O leads to an increased number of metabolite features detected and increased metabolome coverage. Specifically, the use of pure water for sample reconstitution increases the number of significant features identified in human serum samples; consequently, 100 % H_2O increases the chances of discovering novel blood-based biomarkers.

2.4 GENERAL CONCLUSIONS AND FUTURE PERSPECTIVES

When the use of untargeted quantitative metabolomics became established, the main focus was on biomarker discovery applications. In comparison to other omics techniques, metabolomics provides a direct read-out of the active cellular phenotype which makes it a research area with large potential. However, without denying the contribution of metabolomics to biomarker discovery, over time it has become clear that a quantitative metabolic profile only does not provide enough information. It should be seen above all as a starting point for hypothesis generation where unexpected changes in metabolite levels can be exploited to uncover novel mechanistic knowledge in health as well as disease. The four studies included in this thesis mirror this development in the field, starting from paper I. Here, we performed a classic metabolite biomarker study, validating our results in a second, independent cohort. In paper II, we discussed the inclusion of control sample groups in biomarker studies to improve the chances of identifying disease-specific metabolite markers. In paper IV (which from this point of view should have been paper III) we evaluated the impact of a sample preparation step on metabolome coverage in untargeted metabolomics studies; to increase the potential of uncovering novel perturbations in metabolite levels, the metabolome coverage should be maximized. Finally, in paper III, we applied untargeted metabolomics, as well as mRNA sequencing, to generate hypotheses for the mechanisms of drug-resistance in cancer cells. We then proceeded to further investigate these hypotheses using complementary methods such as viability assays and stable isotope tracing experiments, in an effort to increase our understanding of the connection between cell metabolism and drug-resistance. Once again, it needs to be pointed out that untargeted metabolomics should not be seen as a stand-alone analytical technique. Instead, metabolomics discovery results should be incorporated in a larger biochemical context by mechanistic investigations performed by the use of complementary techniques.

3 ACKNOWLEDGEMENTS

Anders, det är svårt att veta vad jag ska säga. Men egentligen är det nog ganska enkelt: tack Anders. Du är en otroligt inspirerande och kreativ forskare. Det var din entusiasm som fick mig att vilja doktorera, och utan dig hade den här avhandlingen aldrig blivit färdig. Du vet väl att du alltid är välkommen förbi på dillkött?

Thank you dear co-supervisors! **Janne**, tack för att du välkomnade mig i din grupp, och tack för ditt stöd och kloka ord när de behövdes som mest. Thanks **Rainer** and **Roland** for interesting discussions.

Hillevi, utan våra tidiga morgonsamtal över valfri varm dyck hade jag aldrig klarat det här, och det menar jag verkligen. Du är rolig, smart och klok. Grönt är den vackraste färgen!

Helena, du lärde mig att se de enskila träden och då blev skogen inte lika mörk och ogenomtränglig längre (ursäkta metaforen man jag kunde inte låta bli!).

Tack **Jenny** för din knivskarpa förmåga att skala bort allt oväsentligt och sedan sy ihop säcken ;-) Och lika stort tack för all din smittande optimism och energi, jag blir alltid glad av att träffa dig!

Hasan, you are a truly kind person. The lab felt just a little bit colder when you left. I wish you all the best for the future!

Thanks to everyone in the **Lehtiö Lab Clinical Proteomics MS group**, past and present members, for creating such a nice atmosphere. **Maria**, **Ann-Sofi**, **Jessie**, **Hanna** och **Lina**, det är tomt utan er på kontoret. **Jorrit**, att sitta bredvid dig är det enda jag saknar med kontorslandskapet. Du vet hur man bär upp en rosa t-shirt! **Henrik** and **Rui**, thanks for putting up with me running my dirty samples on the LTQ Orbitrap Velos back in the day ;-)

Thanks to the **Nordström group** in Umeå, always making me feel welcome: **Claudia**, **Siv**, **Oliver** and **Daniel**. Och **Johan Kolmert**. För att du har lärt mig att problemlösning inte ska göras ensam!

Tack till **Swedish Metabolomics Centre**, särskilt till **Maria Ahnlund**, **Jonas Gullberg** och **Annika Johansson**, för utmärkt service och trevligt bemötande.

Axel Mie, alltid lika trevligt att träffas och äta lunch. Hoppas att vi kommer att kunna jobba ihop någon gång igen!

Kerstin Röhne, att du läser det här betyder att du har den fysiska boken i din hand. Äntligen! Tack för att du hjälpte mig att komma hit, nu kan jag dansa igen!

Anneli och **Andrea** och **Lovisa**, för att ni har lyssnat, stöttat och uppmuntrat. Lovisa utan dina Stenmarkare ingen avhandling. Japan nästa vår?

Tack min allra käraste familj. Tack **Matti** och **Lars**, jag är så glad att ni är mina föräldrar! Tack för alla onsdagar och fredagar kära söstra mi **Malin**. Tack **Sebastian**, **Amelie** och **Lilian** för att vi alltid är välkomna hemma hos er.

4 REFERENCES

Altman, B. J., Z. E. Stine and C. V. Dang (2016). "From Krebs to clinic: glutamine metabolism to cancer therapy." Nature Reviews Cancer **16**(10): 619-634.

Anders, S. and W. Huber (2010). "Differential expression analysis for sequence count data." Genome Biology **11**(10).

Andronesi, O. C., O. Rapalino, E. Gerstner, A. Chi, T. T. Batchelor, D. P. Cahil, A. G. Sorensen and B. R. Rosen (2013). "Detection of oncogenic IDH1 mutations using magnetic resonance spectroscopy of 2-hydroxyglutarate." Journal of Clinical Investigation **123**(9): 3659-3663.

Buscher, J. M., D. Czernik, J. C. Ewald, U. Sauer and N. Zamboni (2009). "Cross-platform comparison of methods for quantitative metabolomics of primary metabolism." Anal Chem **81**(6): 2135-2143.

Cech, N. B. and C. G. Enke (2001). "Practical implications of some recent studies in electrospray ionization fundamentals." Mass Spectrometry Reviews **20**(6): 362-387.

Chechlinska, M., M. Kowalewska and R. Nowak (2010). "Systemic inflammation as a confounding factor in cancer biomarker discovery and validation." Nature Reviews Cancer **10**(1): 2-U13.

Dalgliesh, C. E., E. C. Horning, M. G. Horning, K. L. Knox and K. Yarger (1966). "A GAS-LIQUID-CHROMATOGRAPHIC PROCEDURE FOR SEPARATING A WIDE RANGE OF METABOLITES OCCURRING IN URINE OR TISSUE EXTRACTS." Biochemical Journal **101**(3): 792-+.

Drogan, D., W. B. Dunn, W. Lin, B. Buijsse, M. B. Schulze, C. Langenberg, M. Brown, A. Floegel, S. Dietrich, O. Rolandsson, D. C. Wedge, R. Goodacre, N. G. Forouhi, S. J. Sharp, J. Spranger, N. J. Wareham and H. Boeing (2015). "Untargeted metabolic profiling identifies altered serum metabolites of type 2 diabetes mellitus in a prospective, nested case control study." Clin Chem **61**(3): 487-497.

Dunn, W. B., D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J. D. Knowles, A. Halsall, J. N. Haselden, A. W. Nicholls, I. D. Wilson, D. B. Kell, R. Goodacre and H. C. Human Serum Metabolome (2011). "Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry." Nature Protocols **6**(7): 1060-1083.

Dunn, W. B., W. Lin, D. Broadhurst, P. Begley, M. Brown, E. Zelena, A. A. Vaughan, A. Halsall, N. Harding, J. D. Knowles, S. Francis-McIntyre, A. Tseng, D. I. Ellis, S. O'Hagan, G. Aarons, B. Benjamin, S. Chew-Graham, C. Moseley, P. Potter, C. L. Winder, C. Potts, P. Thornton, C. McWhirter, M. Zubair, M. Pan, A. Burns, J. K. Cruickshank, G. C. Jayson, N. Purandare, F. C. Wu, J. D. Finn, J. N. Haselden, A. W. Nicholls, I. D. Wilson, R. Goodacre and D. B. Kell (2015). "Molecular phenotyping of a UK population: defining the human serum metabolome." Metabolomics **11**: 9-26.

Fang, M., J. Ivanisevic, H. P. Benton, C. H. Johnson, G. J. Patti, L. T. Hoang, W. Uritboonthai, M. E. Kurczy and G. Siuzdak (2015). "Thermal Degradation of Small Molecules: A Global Metabolomic Investigation." Anal Chem **87**(21): 10935-10941.

Fang, M. L., J. Ivanisevic, H. P. Benton, C. H. Johnson, G. J. Patti, L. T. Hoang, W. Uritboonthai, M. E. Kurczy and G. Siuzdak (2015). "Thermal Degradation of Small Molecules: A Global Metabolomic Investigation." Analytical Chemistry **87**(21): 10935-10941.

Fiehn, O., J. Kopka, P. Dormann, T. Altmann, R. N. Trethewey and L. Willmitzer (2000). "Metabolite profiling for plant functional genomics." Nature Biotechnology **18**(11): 1157-1161.

Furey, A., M. Moriarty, V. Bane, B. Kinsella and M. Lehane (2013). "Ion suppression; A critical review on causes, evaluation, prevention and applications." Talanta **115**: 104-122.

Galluzzi, L., O. Kepp, M. G. Vander Heiden and G. Kroemer (2013). "Metabolic targets for cancer therapy." Nature Reviews Drug Discovery **12**(11): 829-846.

Gonzalez-Dominguez, R., T. Garcia-Barrera and J. L. Gomez-Ariza (2015). "Metabolite profiling for the identification of altered metabolic pathways in Alzheimer's disease." Journal of Pharmaceutical and Biomedical Analysis **107**: 75-81.

Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.

Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of Cancer: The Next Generation." Cell **144**(5): 646-674.

Hemstrom, P. and K. Irgum (2006). "Hydrophilic interaction chromatography." Journal of Separation Science **29**(12): 1784-1821.
Horai, H., M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito and T. Nishioka (2010). "MassBank: a public repository for sharing mass spectral data for life sciences." Journal of Mass Spectrometry **45**(7): 703-714.
Horning, E. C. and M. G. Horning (1971). "METABOLIC PROFILES - GAS-PHASE METHODS FOR ANALYSIS OF METABOLITES." Clinical Chemistry **17**(8): 802-&.
Ivanisevic, J., Z. J. Zhu, L. Plate, R. Tautenhahn, S. Chen, P. J. O'Brien, C. H. Johnson, M. A. Marletta, G. J. Patti and G. Siuzdak (2013). "Toward 'omic scale metabolite profiling: a dual separation-mass spectrometry approach for coverage of lipid and central carbon metabolism." Anal Chem **85**(14): 6876-6884.
Jentzmik, F., C. Stephan, M. Lein, K. Miller, B. Kamlage, B. Bethan, G. Kristiansen and K. Jung (2011). "Sarcosine in Prostate Cancer Tissue is Not a Differential Metabolite for Prostate Cancer Aggressiveness and Biochemical Progression." Journal of Urology **185**(2): 706-711.
Jentzmik, F., C. Stephan, K. Miller, M. Schrader, A. Erbersdobler, G. Kristiansen, M. Lein and K. Jung (2010). "Sarcosine in Urine after Digital Rectal Examination Fails as a Marker in Prostate Cancer Detection and Identification of Aggressive Tumours." European Urology **58**(1): 12-18.
Johnson, C. H., J. Ivanisevic and G. Siuzdak (2016). "Metabolomics: beyond biomarkers and towards mechanisms." Nature Reviews Molecular Cell Biology **17**(7): 451-459.
Koeth, R. A., Z. E. Wang, B. S. Levison, J. A. Buffa, E. Org, B. T. Sheehy, E. B. Britt, X. M. Fu, Y. P. Wu, L. Li, J. D. Smith, J. A. DiDonato, J. Chen, H. Z. Li, G. D. Wu, J. D. Lewis, M. Warrier, J. M. Brown, R. M. Krauss, W. H. W. Tang, F. D. Bushman, A. J. Lusis and S. L. Hazen (2013). "Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis." Nature Medicine **19**(5): 576-585.
Kowalewska, M., R. Nowak and M. Chechlinska (2010). "Implications of cancer-associated systemic inflammation for biomarker studies." Biochimica Et Biophysica Acta-Reviews on Cancer **1806**(2): 163-171.
Kuhl, C., R. Tautenhahn, C. Bottcher, T. R. Larson and S. Neumann (2012). "CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets." Analytical Chemistry **84**(1): 283-289.
Markley, J. L., R. Bruschweiler, A. S. Edison, H. R. Eghbalnia, R. Powers, D. Raftery and D. S. Wishart (2016). "The future of NMR-based metabolomics." Curr Opin Biotechnol **43**: 34-40.
Mayerle, J., H. Kalthoff, R. Reszka, B. Kamlage, E. Peter, B. Schniewind, S. Gonzalez Maldonado, C. Pilarsky, C. D. Heidecke, P. Schatz, M. Distler, J. A. Scheiber, U. M. Mahajan, F. U. Weiss, R. Grutzmann and M. M. Lerch (2017). "Metabolic biomarker signature to differentiate pancreatic ductal adenocarcinoma from chronic pancreatitis." Gut.
Mayers, J. R., C. Wu, C. B. Clish, P. Kraft, M. E. Torrence, B. P. Fiske, C. Yuan, Y. Bao, M. K. Townsend, S. S. Tworoger, S. M. Davidson, T. Papagiannakopoulos, A. Yang, T. L. Dayton, S. Ogino, M. J. Stampfer, E. L. Giovannucci, Z. R. Qian, D. A. Rubinson, J. Ma, H. D. Sesso, J. M. Gaziano, B. B. Cochrane, S. M. Liu, J. Wactawski-Wende, J. E. Manson, M. N. Pollak, A. C. Kimmelman, A. Souza, K. Pierce, T. J. Wang, R. E. Gerszten, C. S. Fuchs, M. G. Vander Heiden and B. M. Wolpin (2014). "Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development." Nature Medicine **20**(10): 1193-1198.
Mehlem, A., I. Palombo, X. Wang, C. E. Hagberg, U. Eriksson and A. Falkevall (2016). "PGC-1 alpha Coordinates Mitochondrial Respiratory Capacity and Muscular Fatty Acid Uptake via Regulation of VEGF-B." Diabetes **65**(4): 861-873.
Nordstrom, A. and R. Lewensohn (2010). "Metabolomics: moving to the clinic." J Neuroimmune Pharmacol **5**(1): 4-17.
Nordstrom, A., E. Want, T. Northen, J. Lehtio and G. Siuzdak (2008). "Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics." Analytical Chemistry **80**(2): 421-429.

Patti, G. J., O. Yanes and G. Siuzdak (2012). "Innovation: Metabolomics: the apogee of the omics trilogy." Nat Rev Mol Cell Biol **13**(4): 263-269.
Pauling, L., A. B. Robinson, Teranish.R and P. Cary (1971). "QUANTITATIVE ANALYSIS OF URINE VAPOR AND BREATH BY GAS-LIQUID PARTITION CHROMATOGRAPHY." Proceedings of the National Academy of Sciences of the United States of America **68**(10): 2374-&.
Pavlova, N. N. and C. B. Thompson (2016). "The Emerging Hallmarks of Cancer Metabolism." Cell Metabolism **23**(1): 27-47.
Petrak, J., R. Ivanek, O. Toman, R. Cmejla, J. Cmejlova, D. Vyoral, J. Zivny and C. D. Vulpe (2008). "Deja vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins." Proteomics **8**(9): 1744-1749.
Sevastou, I., E. Kaffe, M. A. Mouratis and V. Aidinis (2013). "Lysoglycerophospholipids in chronic inflammatory disorders: The PLA(2)/LPC and ATX/LPA axes." Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids **1831**(1): 42-60.
Shi, L., C. Brunius, M. Lindelof, S. A. Shameh, H. Wu, I. Lee, R. Landberg and A. A. Moazzami (2016). "Targeted metabolomics reveals differences in the extended postprandial plasma metabolome of healthy subjects after intake of whole-grain rye porridges versus refined wheat bread." Mol Nutr Food Res.
Smith, C. A., G. O'Maille, E. J. Want, C. Qin, S. A. Trauger, T. R. Brandon, D. E. Custodio, R. Abagyan and G. Siuzdak (2005). "METLIN - A metabolite mass spectral database." Therapeutic Drug Monitoring **27**(6): 747-751.
Smith, C. A., E. J. Want, G. O'Maille, R. Abagyan and G. Siuzdak (2006). "XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification." Anal Chem **78**(3): 779-787.
Sreekumar, A., L. M. Poisson, T. M. Rajendiran, A. P. Khan, Q. Cao, J. D. Yu, B. Laxman, R. Mehra, R. J. Lonigro, Y. Li, M. K. Nyati, A. Ahsan, S. Kalyana-Sundaram, B. Han, X. H. Cao, J. Byun, G. S. Omenn, D. Ghosh, S. Pennathur, D. C. Alexander, A. Berger, J. R. Shuster, J. T. Wei, S. Varambally, C. Beecher and A. M. Chinnaiyan (2009). "Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression." Nature **457**(7231): 910-914.
Staubert, C., R. Krakowsky, H. Bhuiyan, B. Witek, A. Lindahl, O. Broom and A. Nordstrom (2016). "Increased lanosterol turnover: a metabolic burden for daunorubicin-resistant leukemia cells." Medical Oncology **33**(1).
Suhre, K. and C. Gieger (2012). "Genetic variation in metabolic phenotypes: study designs and applications." Nature Reviews Genetics **13**(11): 759-769.
Suhre, K., S. Y. Shin, A. K. Petersen, R. P. Mohney, D. Meredith, B. Wagele, E. Altmaier, P. Deloukas, J. Erdmann, E. Grundberg, C. J. Hammond, M. H. de Angelis, G. Kastenmuller, A. Kottgen, F. Kronenberg, M. Mangino, C. Meisinger, T. Meitinger, H. W. Mewes, M. V. Milburn, C. Prehn, J. Raffler, J. S. Ried, W. Romisch-Margl, N. J. Samani, K. S. Small, H. E. Wichmann, G. J. Zhai, T. Illig, T. D. Spector, J. Adamski, N. Soranzo, C. Gieger and CardioGram (2011). "Human metabolic individuality in biomedical and pharmaceutical research." Nature **477**(7362): 54-U60.
Thiele, I., N. Swainston, R. M. T. Fleming, A. Hoppe, S. Sahoo, M. K. Aurich, H. Haraldsdottir, M. L. Mo, O. Rolfsson, M. D. Stobbe, S. G. Thorleifsson, R. Agren, C. Bolling, S. Bordel, A. K. Chavali, P. Dobson, W. B. Dunn, L. Endler, D. Hala, M. Hucka, D. Hull, D. Jameson, N. Jamshidi, J. J. Jonsson, N. Juty, S. Keating, I. Nookaew, N. Le Novere, N. Malys, A. Mazein, J. A. Papin, N. D. Price, E. Selkov, M. I. Sigurdsson, E. Simeonidis, N. Sonnenschein, K. Smallbone, A. Sorokin, J. van Beek, D. Weichart, I. Goryanin, J. Nielsen, H. V. Westerhoff, D. B. Kell, P. Mendes and B. O. Palsson (2013). "A community-driven global reconstruction of human metabolism." Nature Biotechnology **31**(5): 419-+.
Trezzi, J. P., N. Vlassis and K. Hiller (2015). The Role of Metabolomics in the Study of Cancer Biomarkers and in the Development of Diagnostic Tools. Advances in Cancer Biomarkers: From Biochemistry to Clinic for a Critical Revision. R. Scatena. **867:** 41-57.
Trygg, J. and S. Wold (2002). "Orthogonal projections to latent structures (O-PLS)." Journal of Chemometrics **16**(3): 119-128.
Vander Heiden, M. G. and R. J. DeBerardinis (2017). "Understanding the Intersections between Metabolism and Cancer Biology." Cell **168**(4): 657-669.

Vaniya, A. and O. Fiehn (2015). "Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics." Trac-Trends in Analytical Chemistry **69**: 52-61.
Wang, T. J., M. G. Larson, R. S. Vasan, S. Cheng, E. P. Rhee, E. McCabe, G. D. Lewis, C. S. Fox, P. F. Jacques, C. Fernandez, C. J. O'Donnell, S. A. Carr, V. K. Mootha, J. C. Florez, A. Souza, O. Melander, C. B. Clish and R. E. Gerszten (2011). "Metabolite profiles and the risk of developing diabetes." Nature Medicine **17**(4): 448-U483.
Want, E. J., P. Masson, F. Michopoulos, I. D. Wilson, G. Theodoridis, R. S. Plumb, J. Shockcor, N. Loftus, E. Holmes and J. K. Nicholson (2013). "Global metabolic profiling of animal and human tissues via UPLC-MS." Nat Protoc **8**(1): 17-32.
Wishart, D. S., T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. F. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. G. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert (2013). "HMDB 3.0-The Human Metabolome Database in 2013." Nucleic Acids Research **41**(D1): D801-D807.
Yin, P. Y. and G. W. Xu (2014). "Current state-of-the-art of nontargeted metabolomics based on liquid chromatography-mass spectrometry with special emphasis in clinical applications." Journal of Chromatography A **1374**: 1-13.
Yu, Z. H., G. Kastenmuller, Y. He, P. Belcredi, G. Moller, C. Prehn, J. Mendes, S. Wahl, W. Roemisch-Margl, U. Ceglarek, A. Polonikov, N. Dahmen, H. Prokisch, L. Xie, Y. X. Li, H. E. Wichmann, A. Peters, F. Kronenberg, K. Suhre, J. Adamski, T. Illig and R. Wang-Sattler (2011). "Differences between Human Plasma and Serum Metabolite Profiles." Plos One **6**(7).

Zamboni, N., A. Saghatelian and G. J. Patti (2015). "Defining the Metabolome: Size, Flux, and Regulation." Molecular Cell **58**(4): 699-706.