From the Research Center Borstel

Leibniz Center for Medicine and Biosciences

Director: Prof. Dr. Stefan Ehlers

Division of Bioanalytical Chemistry

Head: Dr. Dominik Schwudke

Systematic investigation of lipid profiles from human lung tissues reveals specific lipid*ome* alterations in lung cancer and pulmonary emphysema

Dissertation

for Fulfillment of

Requirements

for the Doctoral Degree

of the University of Lübeck

from the Department of Natural Sciences

Submitted by

Lars Florian Eggers

from Bad Oldesloe

Lübeck 2017

First referee: Prof. Dr. Torsten Goldmann

Second referee: Prof. Dr. Lars Redecke

Date of oral examination: January 31, 2018

Approved for printing. Lübeck, Febuary 5, 2018

Abstract

Little is known about the lipid metabolism of human lungs and the interplay of lipid metabolic processes in the development of lung cancer and chronic obstructive pulmonary disease (COPD). Within the past decade, numerous studies discovered lipid metabolic perturbations in diseases including cardiovascular diseases and cancer [1] and revealed the impact of lipid metabolic processes in pathogenesis. For instance, specific lipid biomarkers were identified for diagnosis and prediction of disease progression in ovarian and lung cancer [2–4]. Other studies showed specific lipid*ome* alterations in induced sputum of COPD patients with respect to their ceramide levels [5]. However, no studies were carried out which provided a comprehensive characterization of the lipid composition (lipid*ome*) in human lungs. Furthermore, no investigation was carried out to determine how the heterogeneous morphological structure of tissue samples is reflected in their lipid*omes*.

This dissertation provides a systematic study of the lipid*omes* from alveolar human lung tissues and from lung cancer tissues. One hundred seventy-four tissue samples were collected and analyzed from lung cancer patients after surgical tumor removal. *Top-down* lipid*omics* screens were performed and enabled quantification of about 400 lipid species from 14 classes.

This study identified widespread alterations in the lipid profiles between tumor-free alveolar tissues and the corresponding tumors. Lipid components of the *pulmonary surfactant*, which is a lipid-protein complex lining the alveoli, such as phosphatidylglycerol and saturated phosphatidylcholine (PC) were present in high levels in alveolar tissues. Increased amounts of the main constituents of cellular membranes, including PC 34:1 and cholesterol, were shown in tumor tissues. A selected set of surfactant and cell membrane-associated lipids was sufficient to distinguish between alveolar and tumor tissues with high diagnostic confidence. This study revealed specific alterations between the lipid*omes* of the major entities of non-small-cell lung cancer (NSCLC), adenocarcinoma (ADC) and squamous-cell carcinoma (SCC). The lipid*omes* of carcinoid tumors, which are slowly growing and low malignant tumors, differed significantly from the NSCLC lipid*omes*. In particular, the carcinoids had altered ceramide levels. Even lipid*omes* of the tumor-free alveolar tissues of ADC and SCC patients had distinct features within their lipid profiles that made them distinguishable from each other.

This work demonstrated that a modeling approach based on partial least-squares regression

is suitable to provide dependencies of the tissue lipid*ome* composition from the respective histopathological phenotype. Moreover, the data revealed interplay of triacylglycerols with necrotic areas within the tissues. Additionally, this study demonstrated how an altering tissue composition regarding the proportions of metabolically active tumor cells, stroma and necrosis, associated with the tissue lipid*omes* of tumors. In the tumor-free alveolar lung tissue samples, the lipid profiles changed with the age of the patients. The lipid*omes* mirrored a correlation of aging with the progression of the pulmonary emphysema, which is a pathological and irreversible enlargement of the alveolar structure and a common phenotype of COPD. Furthermore, the results showed that lipid*omes* from alveolar lung tissues differed between the genders.

This dissertation provides first insight into lipid metabolic processes of lung tissues. It delivers a comprehensive description of a lipid metabolic snapshot including the major abundant glycerolipids, glycerophospholipids, sphingolipids and cholesterols. For the first time, lipid metabolic alterations induced by the heterogeneous histopathological phenotype demonstrates connections between quantitative lipid*omes* and the tissue morphology. This opens possibilities for discovery of specific diagnostic markers and lipid metabolic pathways relevant for carcinogenesis and COPD. This study provides a comprehensive lipid metabolic snapshot of human lung tissues and this approach could be the basis for thorough investigation of metabolic pathways to understand the underlying mechanisms in pathogenesis.

Zusammenfassung

Lipidmetabolische Prozesse der humanen Lunge sind derzeit nur wenig erforscht. Insbesondere über den Einfluss dieser Prozesse auf die Entwicklung von Lungenkrebs oder chronisch-obstruktiver Lungenerkrankung (COPD) liegen bislang nur wenige Daten vor. Untersuchungen lipidmetabolischer Prozesse, die innerhalb der letzten zehn Jahre durchgeführt wurden, zeigten spezifische Veränderungen in Herz-Kreislauf- und Krebserkrankungen [1] und damit, wie groß der Einfluss des Lipidstoffwechsels in der Pathogenese sein kann. Dabei konnten spezifische Lipid-Biomarker zur Diagnose und zur Vorhersage der Krankheitsverläufe von Eierstockkrebs und Lungenkrebs identifiziert werden [2–4]. Eine andere Studie zeigte spezifische Veränderungen der Ceramidspiegel im induzierten Sputum von COPD-Patienten [5]. Es liegen aber keine Studien vor, die eine umfassende Charakterisierung der Lipidzusammensetzung (Lipid*ome*) von Lungengeweben liefern. Darüber hinaus gibt es kaum Untersuchungen, die belegen, inwiefern sich die sehr heterogene morphologische Struktur von Gewebeproben auf ihre Lipid*ome* auswirkt.

In dieser Dissertation wurden die Lipid*ome* von humanen tumorfreien alveolären Lungengeweben und von Lungenkarzinomen charakterisiert und quantitativ beschrieben. Dazu wurden 174 Gewebeproben von 92 Lungenkrebspatienten untersucht. Nach chirurgischer Entfernung des Lungentumors wurden *top-down* Lipid*omics*-Screens durchgeführt, womit etwa 400 Lipide aus 14 Lipidklassen stabil und reproduzierbar quantifiziert wurden. Parallel dazu wurden die selben Gewebeproben histopathologisch hinsichtlich ihrer Morphologie charakterisiert und ihr spezifischer Phänotyp in quantitativen Scores ausgedrückt.

Die Lipidzusammensetzungen der Tumorgewebeproben zeigten charakteristische Veränderungen im Vergleich zu den tumorfreien alveolären Lungengewebeproben. Die Lipid-Bestandteile des Lungen*surfactants*, ein die Alveolen benetzender Lipid-Protein-Komplex, wie beispielsweise Phosphatidylglycerin und gesättigtes Phosphatidylcholin (PC), wurden in hohen Konzentrationen in den alveolären Lungengeweben wiedergefunden. Lipide, die Hauptbestandteile zellulärer Membranen sind, wie beispielsweise PC 34:1 und freies Cholesterin, wurden in erhöhten Konzentrationen in den Tumorgeweben nachgewiesen. Anhand der charakteristischen Lipid*om*veränderungen zwischen Alveolär- und Tumorgeweben, wurden repräsentative Lipide ausgewählt, die Tumorgewebe von den alveolären Geweben eindeutig unterschieden.

Es wurden Lipidmarker identifiziert, die es möglich machten, die Hauptentitäten des nichtkleinzelligen Lungenkarzinoms (NSCLC), Adenokarzinoms (ADC) und Plattenepithelkarzinoms (SCC), voneinander zu unterscheiden. Insbesondere karzinoide Tumorgewebe, bei welchen es sich meist um langsam wachsende und niedrig maligne Tumoren handelt, unterschieden sich von den NSCLC-Geweben in ihren Lipidprofilen, was sich am deutlichsten in ihren Ceramid-Konzentrationen widerspiegelte. Sogar die Lipid*ome* des tumor-freien alveolären Lungengewebes unterschieden sich in ADC und SCC Patienten.

Diese Arbeit zeigt, dass ein Modellierungsansatz der auf Partial-Least-Squares-Regressionen beruht, spezifische Einflüsse der histopathologischen und morphologischen Eigenschaften auf die Lipid*ome* der Gewebeproben offenlegt. So gab es beispielsweise Hinweise darauf, dass erhöhte Triacylglycerine spezifisch mit nekrotischen Regionen der Tumorgewebeproben assoziieren. In den tumorfreien, alveolären Lungengewebeproben wurde ein klarer Zusammenhang zwischen den Lipidprofilen und dem Patientenalter hergestellt. Dieser Parameter korrelierte zudem mit dem Grad des Lungen*emphysems*, welches eine pathologische und irreversible Erweiterung der Alveolärstruktur darstellt und ein häufiger Phänotyp der COPD ist. Es konnten sogar spezifische Einflüsse des Geschlechtes der Patienten auf ihre Lipidprofile gefunden werden.

Mit diesen Ergebnissen liefert die vorliegende Dissertation erste Einblicke in die Lipidzusammensetzung der humanen Lunge. Somit zeigt diese Arbeit zum ersten Mal einen Zusammenhang zwischen der quantitativen Lipid*om*zusammensetzung und dem histopathologischen Phänotyp der Gewebeproben. Diese Methodik eröffnet neue Möglichkeiten, hoch-spezifische diagnostische Marker zu entdecken und außerdem Lipid-Stoffwechselwege zu beschreiben, die für die Karzinogenese und COPD relevant sind. Hiermit liefert diese Studie einen umfassenden Überblick über den lipidmetabolischen Zustand von humanen Lungengeweben. Dieser Ansatz könnte die Grundlage für tiefergehende Untersuchungen lipidmetabolischer Routen bilden, um die grundlegenden Mechanismen der Pathogenese zu verstehen.

Contents

1 Introduction

1.1 Lipid*omics* enable association of the lipid metabolism with disease progression

This dissertation deals with the interplay between lipid profiles in human lung tissues and lung pathologies such as *e. g.* lung cancer or pulmonary emphysema. Lipids have crucial biological functions. They are major components of cellular membranes, store chemical energy and are mediators for cellular signaling [6, 7]. Although good extraction protocols for lipids are available since the 1950s [8, 9], thorough investigations of lipid metabolic processes have been difficult for a long time due to time-consuming analysis techniques. This changed at the beginning of the 21st century. The advances of modern mass spectrometric techniques over the past decade allowed the rapid development of lipid*omics* that deals with the quantification of a most comprehensive lipid metabolic snapshot within a biological system: the lipid*ome*. The modern analytical platforms used in lipid*omics* enable rapid identification and quantification of hundreds of lipid species, including fatty acids (FAs), glycerolipids (GLs), glycerophospholipids (GPLs), sphingolipids (SLs) and cholesterols [10–14]. These newly developed analytical approaches allow researchers to investigate lipid metabolic processes in disease development. For example, disorders in lipid metabolic processes have been associated with various disease states including Alzheimer's disease [15], schizophrenia [16], obesity [17, 18], type II diabetes [19], cardiovascular disease, and cancer [1]. Studies discovered lipid*ome* alterations associated with respiratory diseases. For instance, Ollero *et al.* [20] discovered lipid signatures in the plasma of cystic fibrosis patients, Telenga *et al.* [5] showed SL alterations in the induced sputum of chronic obstructive pulmonary disease (COPD) patients and Kang *et al.* [21] analyzed brochio-alveolar lavage fluid (BALF) of asthma patients. Moreover, Hall *et al.* [22] revealed

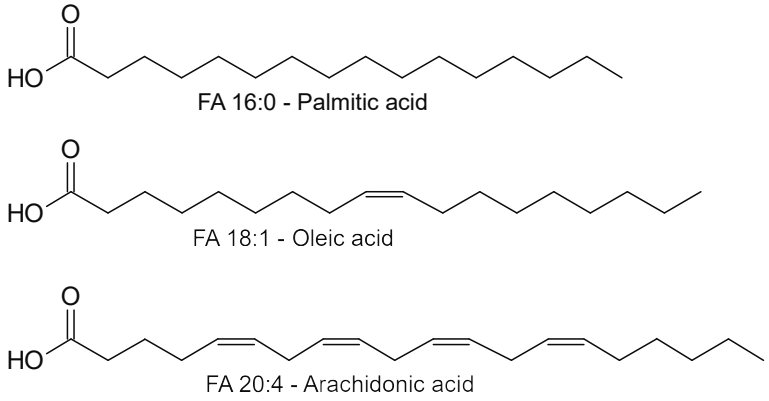


Figure 1.1: **Chemical structures of common fatty acids.**

specific lipid metabolic perturbations in a lung cancer mouse model. However, lipid*omics* is a relatively new discipline and there are still many open questions. For example, the role of lipid metabolic processes during carcinogenesis is not fully understood, as well as the influence of the highly specialized lipid metabolism in diseases of the respiratory tract such as the development of emphysema in COPD.

1.2 Lungs have a highly specialized lipid metabolism

Lipids play crucial roles in human lungs and their classes and functions in human lungs as well as in carcinogenic processes are introduced. Lipid molecules, especially GPL species, have surface-active properties due to their amphiphilic nature. This allows constitution of cellular membranes, vesicles and lipid droplets. These surface-active properties are essential in the lung, not only since they form highly specialized membranes of the apical cell layers. As *surface-active agent*s (surfactants), lipids regulate surface tension at the air/water interface of the lung during breathing. Without this regulation, the alveoli would collapse [23].

A large number of lipid classes are known, each with unique properties and functions. Therefore, a brief introduction to the high variability of lipid classes, their structures, their biosynthesis, and their functions is given below. Due to their anatomical organization, the lungs maintain a large interface to the environment and have a highly specialized epithelial cell layer to protect this interface. From this perspective, it is necessary to further introduce the anatomy of the lung and in which manner the highly specialized lipid metabolic processes are involved in lung homeostasis.

1.2.1 Structures and properties of lipids in the lung

Fatty acids. FAs are carboxylic acids having an aliphatic chain usually containing 14 to 22 carbon atoms. Even FAs with longer carbon chains are known but have low concentrations in biological systems [24]. The aliphatic chains may be saturated, *e. g.* for palmitic acid, abbreviated as FA 16:0, which stands for 16 carbon atoms in the aliphatic chain and 0 carbon-carbon double bonds (dbs). In addition, FAs may also be unsaturated such as oleic acid (FA 18:1). An example for poly-unsaturated fatty acids (PUFAs) is arachidonic acid (FA 20:4) (Fig. 1.1).

FAs can originate either from nutritional sources or can be enzymatically modified or produced by *de-novo* FA synthesis [25]. The eukaryotic FA synthesis and elongation are based on C_2 building blocks, whereby usually fatty acids with even numbers of carbon atoms (even numbered FAs) are synthesized. However, low abundances of fatty acids with odd numbers of carbon atoms (odd numbered FAs) such as C_{15} and C_{17} aliphatic chains are known in humans [26]. These FAs probably originate from nutritional sources [26, 27].

Glycerolipids. GLs have a glycerol backbone as central structural unit which is esterified with FAs (Fig. 1.2, a). The glycerol molecule is prochiral, which means that chirality is induced when the molecule is substituted with different FAs substituents. As a convention, carbon atoms in the glycerol backbone are unambiguously named according to the stereospecific numbering (*sn*) nomenclature [28]. The *sn* nomenclature defines the carbon atom as *sn*-1, which stands top in a Fischer projection in which the substituent of the second carbon is written to the left [28]. The numbering of the glycerol backbone according to the *sn* nomenclature is given in figure 1.2. According to the number of FA substituents, monoacylglycerols (MAGs), diacylglycerols (DAGs) and triacylglycerols (TAGs) are known (Fig. 1.2, a). If there is an excess of free FAs, TAGs are synthesized by the *triglyceride/fatty acid pathway* and can be released *vice versa* [29]. DAG and MAG molecules are released from TAGs or cleaved from GPLs.

Glycerophospholipids. Like GLs, GPLs also have a glycerol backbone as central structural unit. A phosphate group is linked to the glycerol backbone in the *sn*-3 position (Fig. 1.2, b). The phosphate group further acts as linker between the glycerol backbone and a hydrophilic head group (HG) determining the lipid class [30]. Together with the FA substituents, the respective HG modulates the biochemical and biophysical properties of the GPLs and vari-

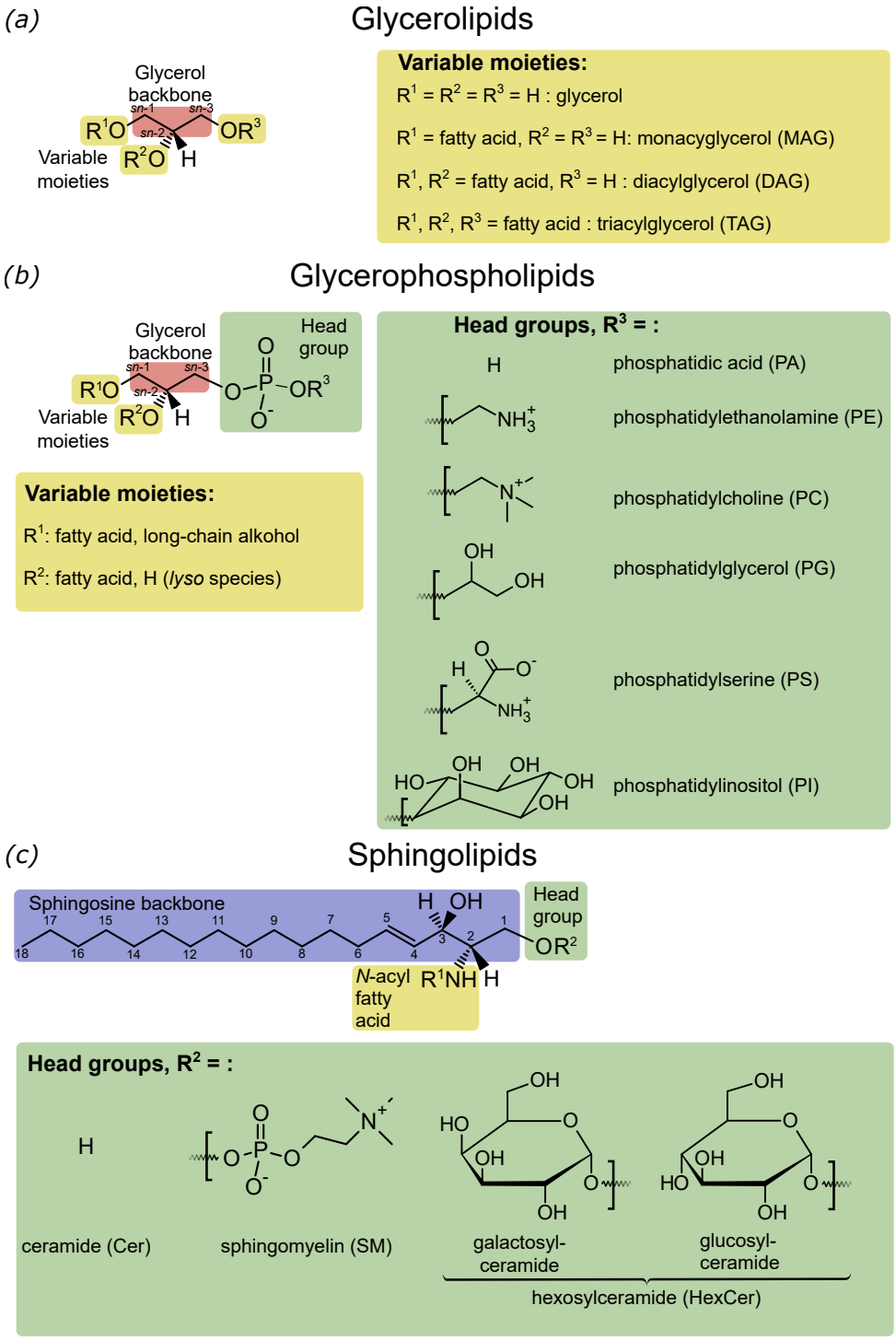


Figure 1.2: **Chemical structures of common lipid classes.** *(a)* Glycerolipids, *(b)* Glycerophospholipids, *(c)* Sphingolipids.

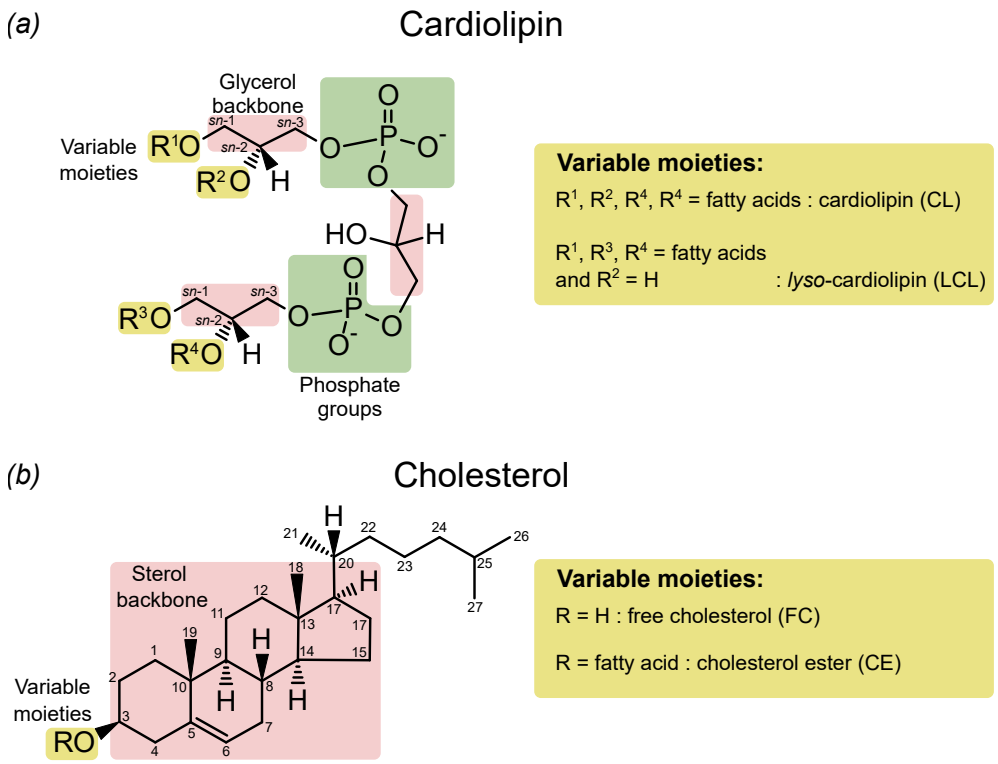


Figure 1.3: **Chemical structures of cardiolipin and cholesterol.**

ous chemical structures are known for them. The structurally simplest GPL is phosphatidic acid (PA), which carries a phosphate as HG (Fig. 1.2, b). The most abundant GPL classes are phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylglycerol (PG), phosphatidylserine (PS) and phosphatidylinositol (PI) in which ethanolamine, choline, glycerol, serine and inositol HGs are linked to the phosphate (Fig. 1.2, b). The *sn*-1 and *sn*-2 positions of the glycerol backbone are linked to FAs which form a hydrophobic tail which allows the lipid molecules to have amphiphilic properties together with the polar HG. This is the basis for their surface-active behavior. Modifications of the unpolar regions of GPLs are known. For instance, *lyso* lipids have a free hydroxylation instead of an acylation in the *sn*-2 position (Fig. 1.2, b). In this case the prefix 'L' is written before the lipid class symbol, *e. g.* *lyso*-phosphatidylcholine (LPC), *lyso*-phosphatidylethanolamine (LPE) and *lyso*-phosphatidic acid (LPA). Instead of an acylation with an FA, the *sn*-1 position in PCs and PEs can also be linked to a long-chain alcohol. These lipids are called *ether lipids* and are labeled with the suffix '*O*-', *e. g.* phosphatidylcholine ether (PC *O*-) and phosphatidylethanolamine ether (PE *O*-).

Plasmalogens are specific cases of ether lipids in which the long-chain alcohol has a *vinyl ether* functional group. The dimeric GPL cardiolipin (CL) consists of two PA molecules linked by an additional glycerol (Fig. 1.3, a).

The central building block for all GPLs is PA, which is synthesized from glycerol-3-phosphate and FAs. DAGs are directly released from PAs, which further act as precursors for the synthesis of PC, PE and TAG. In a separate pathway, cytidine diphosphate (CDP)-DAG is synthesized from PA, which is the biosynthetic precursor for the anionic GPLs such as PG and PI. In mammals, PS is synthesized from PE or PC by an exchange reaction of the respective ethanolamine or choline HGs and serine. There are methyltransferases known, which can convert PE to PC [31]. The synthesis of PG and CL is located in the inner mitochondrial membrane. CDP-DAG and glycerol-3-phosphate are condensed to phosphatidylglycerolphosphate. The cleavage of the terminal phosphate group then releases PG. CDP-DAG and PG react further to CL [31].

Sphingolipids. The long-chain sphingoid base is the central structural unit of SLs [32]. Sphingoid bases have an aliphatic backbone with 14 to 22 carbon atoms [33], are hydroxylated in the 1-position of the aliphatic chain, have a primary amino function in the 2-position, a secondary hydroxylation in the 3-position, and a db in *E*-configuration in the 4-position. The most common sphingoid base is sphingosine, which is based on an aliphatic C_{18} backbone (Fig. 1.2, c). Variations of sphingosine bases are known, such as dihydrosphingosine lacking the db, and phytosphingosine, which is hydroxylated in the 4-position and also has no db [32]. Furthermore, variations in the length of the aliphatic chain are known. The central unit of all SLs is ceramide (Cer) which is a conjunction of the sphingoid base with an FA that is connected as amide with the amino functional group (Fig. 1.2, c). More complex SLs carry specific HGs linked to the primary hydroxylation at the 1-position. The phospho-SL sphingomyelin (SM) has a phosphocholine HG (Fig. 1.2, c). Furthermore, SLs can be glycosylated and are then called glyco-sphingolipids (GSLs). The structurally simplest GSLs have HGs consisting of a glucose or a galactose unit. Since glucose and galactose HGs are not distinguishable by mass spectrometric analysis, they are often summarized as hexosylceramides (HexCers). GSLs can also have complex glycosylation patterns composed of several sugar residues, such as in gangliosides or cerebrosides [34].

SLs are involved in complex metabolic processes. The center of SL metabolism is Cer, which

is synthesized *de-novo* from palmitoyl-coenzyme A (CoA), serine and FAs with dihydroceramide as an intermediate. In subsequent steps, the enzyme sphingomyelinase can transfer a phosphocholine HG to Cer to form SM. Sphingosine and its derivatives such as sphingosine-1-phosphate (S1P) can be cleaved from Cer. Cer can also be phosphorylated to form Cer-1-phosphate or else be glycosylated to form complex GSLs and sulphatides. All these processes are reversible, leading to complex and dynamic SL-metabolic networks that are strictly regulated [35].

Cholesterol. Cholesterols belong to the group of sterol lipids. The cholesterol molecule is based on a steroid structure with a characteristic carbocyclic ring system consisting of three 6-rings and one 5-ring (Fig. 1.3, b). The 3-position of the steroid ring system shows a hydroxylation, in the 5-position a db and in the 17-position a branched C_8 side chain is located (Fig. 1.3, b). The free cholesterol (FC) molecule is hydroxylated in the 3-position and cholesteryl esters (CEs) are acylated with a FA in the 3-position.

The cholesterol biosynthesis follows several steps involving multiple enzymes. It starts with acetate and has *e. g.* 3-hydroxy-3-methylglutaryl CoA and farnesyl pyrophosphate as intermediates. Intermediates with the characteristic steroid backbone are lanosterol and desmosterol [36]. Cholesterol is predominantly synthesized in the liver where it can further act as precursor for bile acids and steroid hormones. For transport, cholesterol is packed into lipoproteins.

1.2.2 Lipid metabolic homeostasis is mandatory for cellular survival

Lipid droplets are a reservoir for neutral lipids. FAs maintain important biological functions since they store biochemical energy and especially the PUFAs are precursors for small lipid mediators such as prostaglandins which are signaling molecules in *e. g.* inflammatory processes and carcinogenesis [37]. Furthermore, FAs are central building blocks of complex lipid structures. TAGs are storage molecules for FAs, which can release free FAs if required. In this manner, TAGs also function as precursors for GPLs and signaling molecules [38]. Similarly, DAGs and MAGs are intermediates in lipid synthesis and precursors of messenger molecules [39]. CE is a storage form of FC. For storage, TAGs condensed together with CEs in lipid droplets [38]. From there, both FC and free FAs can be released [38].

Cellular membranes are highly flexible and adaptable compartments. The introduction into the lipid classes has shown that a large pool of lipids is available in biological systems, each with its own specific properties whereas the function of a single lipid molecule is rather limited, with the exception of signaling processes. The interplay of different lipids allows highly adaptable and flexible structural organization as required for the formation of cellular membranes, which is one of the most important biological function of lipids. Biological membranes are essential for cellular organization. Membranes not only separate the cell interior from the extracellular matrix, they also separate cell organelles such as the mitochondria, the endoplasmic reticulum, the Golgi-apparatus and the nucleus [40]. Each of these organelles has specialized membrane compositions, such as the inner mitochondrial membrane which is enriched with PG and CL. Therefore, the membrane composition is highly adaptive to functional aspects such as fluidity, permeability, curvature and links to membrane associated proteins [41]. Enzymes constantly renew and transform lipids according to actual requirements for instance in repairing processes.

Cellular membranes are predominantly formed by GPLs and FC. The highest proportions of the total amount of GPLs within cell membranes are made of PC molecules with one *Z*-configurated db in their aliphatic chains such as the lipid species PC 34:1. These unsaturated lipids allow fluid membranes under physiological conditions. To achieve adaptive and flexible membrane structures, other GPLs such as PEs, PIs and PS' are incorporated into the membranes [42]. Membrane GPLs may contain PUFAs which are released from the membrane after activation by phospholipases [43]. PI and PS act as messenger molecules besides structural functions in the membrane. FC is essential for cell growth since it is an important component of cell membranes and localizes in the lipophilic part of the phospholipid bilayer where it regulates the membrane fluidity. High levels of the rigid FC molecule in the cell membrane cause an inflexible membrane structure. On the contrary, low amounts of FC in cell membranes lead to flexible and fluid structures [44, 45].

Lipids are not homogenously distributed across the membranes. For example, there are regions with asymmetric distribution of different lipid species across both phospholipid layers to achieve curved membrane structures [40]. Especially in the plasma membrane, FC and SL rich domains, *lipid rafts*, are formed which can bind membrane associated proteins [46]. SLs with complex glycosylation patterns such as gangliosides are responsible for inter-celluar interac-

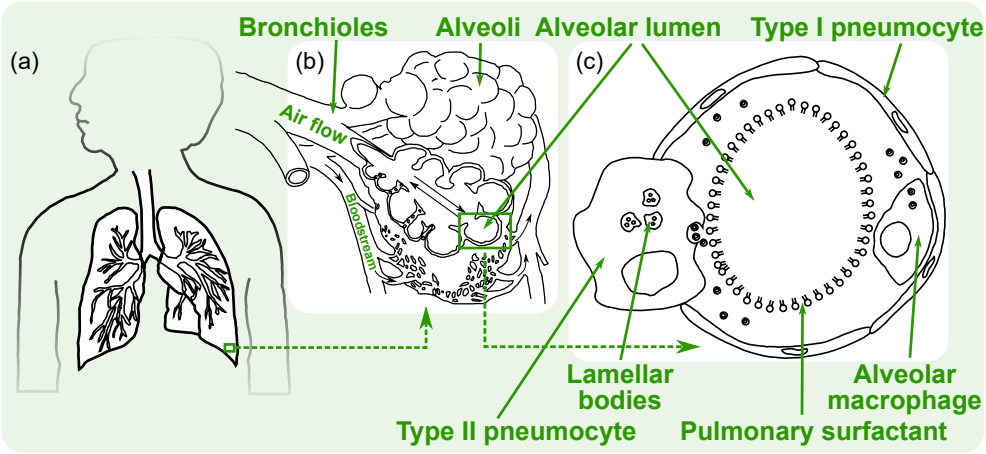


Figure 1.4: **Anatomical organization of human lungs.** *(a)* Structure of the human lung comprising the branching structure of bronchi and bronchioles. *(b)* Organization of alveoli as basic units for gas exchange, bronchioles and blood capillaries. Modified from [49]. *(c)* Cut through an alveolus showing different cell types and pulmonary surfactant lining the inner alveolar surface. Modified from [50].

tions such as cell-contact responses and contact inhibitions when they are presented on the cell surface [47]. These lipid rafts are of particular interest in the lung due to the fact that they are predominantly located in apical domains in the plasma membrane of epithelial cells [46, 48]. From this perspective, elevated amounts of SLs such as SMs are expected in tissues with high proportions of apical membranes. Moreover, SL breakdown products such as sphingosine and S1P act as messenger molecules and were connected with *e. g.* cellular growth and survival, inflammatory processes as well as apoptosis [35]. Cers are known to be involved into signaling processes mediating cell growth, differentiation, necrosis and apoptosis [35].

1.2.3 Physiological processes and lipid metabolism in lung homeostasis

Lipid metabolic processes are essential for lung homeostasis. Lungs are highly specialized organs to allow the exchange of oxygen and carbon dioxide between blood and air [51]. The lung is anatomically structured to cope with a continuous exposure of particles, pathogens and toxins. The air-conducting bronchial tubes have a hierarchical structure. Main bronchi branch into smaller bronchioles, which finally end in the alveoli, the basic units for gas exchange (Fig. 1.4, a). The peripheral regions of the lung contain mainly alveolar structures and tiny bronchioles (Fig. 1.4, b). Therefore, tissue samples from this peripheral area are referred to as *alveolar lung tissues* in this dissertation. The alveolar epithelium covers about 99 % of the surface in

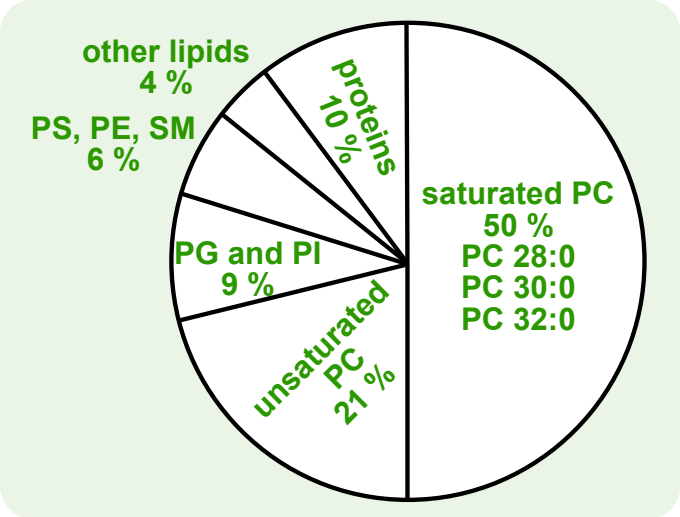


Figure 1.5: **Composition of the pulmonary surfactant.** Modified from [23].

the lung [52] and is in direct contact with the bloodstream, since alveoli are surrounded by blood capillaries. About 95 % of the alveolar surface is covered by alveolar epithelial cells type I (AECI) which have a rod shape and are specialized to form a flat shaped barrier between blood and air of about 1 µm of thickness [52, 53] (Fig. 1.4, c). The remaining 5 % of the alveolar surface is covered by cuboid shaped alveolar epithelial cells type II (AECII). AECII are large cells compared to AECI and characterized by lamellar bodies. Their highly specialized lipid metabolism allows them to synthesize, store and secrete the pulmonary surfactant which is a lipid-protein complex lining the alveolar epithelium.

Pulmonary surfactant regulates surface tension during breathing to prevent alveolar collapse at the end of expiration [23, 51]. The lung surfactant composition is dynamic under environmental exposure and is therefore constantly recycled by AECII as well as by alveolar macrophages (AM) [52]. The pulmonary surfactant comprises a complex mixture of 86 % phospholipids and 10 % surfactant proteins (SPs) A, B, C and D (Fig. 1.5). A fraction of 4 % is composed of neutral lipid species from GL and cholesterol classes which are minor but important surfactant components. SPs fulfill various biochemical functions including structural stabilization and regulation [54]. The hydrophobic SP-B and SP-C are relevant for surfactant film formation and stabilization, while the hydrophilic SP-A and SP-D play critical roles in innate immune responses [23]. The main component of the human lung surfactant is the saturated GPL species PC 16:0/16:0, which is about 50 % of the total surfactant (Fig. 1.5) [55]. Other sat-

urated lipids such as PC 16:0/14:0 are also present in high proportion of pulmonary surfactant. Such saturated lipids are characteristic for the pulmonary surfactant since they are low abundant in cell membranes due to the fact that they tend to form *solid-like* phases in lipid bilayers [42]. Although the overall surfactant composition is conserved among all mammals [23] some mammalian species have another main PC species such as PC 16:0/16:1 in *sminthopsis crassicaudata* (fat-tailed dunnart) [56]. Characteristic for pulmonary surfactant are high levels of GPLs with anionic HGs such as PG and PI which were identified to inhibit inflammatory responses [57]. FC is a component of the lung surfactant and regulates there the fluidity of the surfactant film [58]. PG is a characteristic part of pulmonary surfactant since PG is not present in such high concentrations elsewhere in mammalian systems except in the inner mitochondrial membrane. PG maintains structural as well as immune regulative functions and is therefore important for lipid metabolic homeostasis within the lung [23, 59, 60]. Dysregulations in the surfactant metabolism have been connected to several lung diseases including the respiratory distress syndrome [61] and the interstitial lung disease [62]. It was further shown that bacterial CL causes lung injury during pneumonia [63]. Homeostasis in SL metabolic processes was strongly connected to lung health before. Especially in the lung it was shown that an upregulation of Cers promoted apoptosis and weakened the pulmonary barrier [35, 64]. Moreover, Cer was identified to be involved in lung infections [65].

1.2.4 Lipid metabolic processes of human lungs are altered during carcinogenesis

Lipid metabolic alterations were detected in a variety of cancer diseases [7] including breast cancer [66, 67], kidney cancer [68], prostate cancer [69, 70], colorectal cancer [71, 72], ovarian cancer [73] and lung cancer [74, 75]. In many cases, carcinoma tissues were characterized by increased levels of PC 34:1 compared to the respective normal tissues [66, 68, 73, 74, 76–78] as summarized by Perrotti *et al.* [7].

In recent years, lung cancer came into focus of lipid research. Lung cancer is the leading cause of cancer related deaths [79] and there is a strong need for early stage diagnostic tools as well as treatment strategies. Lung cancer is classified into small-cell lung cancer and non-small-cell lung cancer (NSCLC). Surgical removal of the tumor is the standard therapy for NSCLC.

The most abundant entities of NSCLC are adenocarcinoma (ADC) and squamous-cell carcinoma (SCC). The main cause for NSCLC was associated with the consumption of cigarette smoke. Interestingly, epidemiological studies showed that never-smokers were mainly suffering from lung ADCs [80, 81]. In what way cigarette smoking might influence lipid metabolic processes of human lungs is unclear. However, studies confirmed that the lipid*ome* composition of murine lungs changed after exposure to cigarette smoke [82, 83].

There were studies conducted comparing the lipid profiles from NSCLC tissues and normal lung tissues. Lee *et al.* [74] showed distict lipid profiles of adjacent alveolar lung tissues and tumor tissues revealed by non-quantitative matrix assisted laser desorption/ionization (MALDI)-mass spectrometry (MS). Marien *et al.* [75] identified specific alterations in the quantities of lipids between alveolar and tumor tissues but exclusively focused their study on phospholipids without covering the important classes PG and PA. Both studies reveled that species like PC 34:1, PC 36:2 and some SM species were enriched in NSCLC tissues, where surfactant associated lipids including PC 32:0 species were higher abundant in normal lung tissues [22, 75]. Furthermore, Lee *et al.* and Marien *et al.* showed indications for specific lipid*ome* alterations of ADC and SCC tissues [74, 84].

All these studies clearly indicated specific lipid metabolic alterations in lung cancer. However, a quantitative description of a comprehensive lipid metabolic condition covering the complete spectrum of GLs, GPLs, SLs and cholesterols has not been available today. Furthermore, no approaches were available to measure the quantity of a lipid molecule and its spacial distribution in heterogeneous tissue samples.

1.2.5 Lipids are promising biomarkers in cancer research

The fact that there were lipid metabolic perturbations detectable in several diseases led to studies that investigated lipids as target molecules in biomarker research [7]. Biomarkers are detectable parameters that could provide information about the diagnosis or the progression of certain diseases. These could be proteins such as the prostate specific antigen for prostate cancer [85], genes or metabolites including lipids. A common observation in lipid-biomarker research for cancer was an alteration of *lyso* lipids within the plasma lipid*omes* of cancer patients [86, 87]. For example, LPA molecules were identified as potential biomarkers for ovarian cancer [2, 3].

Furthermore, higher levels of LPC species such as LPC 18:1, LPC 20:3 and LPC 20:4 were quantified in sera from lung cancer patients [4]. Additionally, there were approaches to predict cancer progression from specific lipid markers [88].

For diagnostics, biomarkers should be available from easily accessible sources such as blood, urine or sputum. Earlier studies revealed numerous parameters influencing the human blood plasma lipid*ome* such as aging, gender and body mass index (BMI) [17, 89]. These overlaying effects could hinder the sensitive detection of biomarkers when they are unknown. Discovery of biomarkers is often an empirical process where questions regarding remodeling of metabolic processes are not asked. However, this is mandatory to understand underlying mechanisms of pathogenesis.

1.3 Analysis of lipids from tissues

The previous section showed the importance of lipids in biological functions. Investigation of lipid compositions in tissues and of the underlying biosynthetic pathways is necessary to gain knowledge of the basic mechanisms of pathogenesis. There is a variety of techniques available for selective analysis of the lipid composition in tissues. This section provides an overview of these techniques. It starts with histochemical stains for localization of lipids and enzymes correlated to lipid metabolic processes on tissue sections. Furthermore, mass spectrometric imaging (MSI) is introduced as a method for selective localization of specific molecules on tissue sections. Finally, a quantitative lipid*omics* screening workflow including lipid extraction from biological samples and MS based lipid quantification is described.

1.3.1 Histopathology of tissues and visualization of lipids

Histopathological techniques are capable for detailed morphological analysis of tissue samples. To preserve the tissue morphology and prevent them from degradation, the tissues need to be fixed. Formalin is a routinely used fixative which denaturizes and crosslinks proteins and prevents enzymatic tissue degradation to conserve the samples. However, deoxyribonucleic acids (DNAs), ribonucleic acids (RNAs) and proteins are chemically modified after fixation which hinders analysis of these molecules. Therefore, a fixation method was developed which

is based on the 4-(5-hydroxylethyl)-1-piperazineethanesulfonic acid (HEPES)-glutamic acid buffer mediated organic solvent protection effect (HOPE) [90]. Tissues prepared with the HOPE technique can be used later *e. g.* to extract RNA and to perform transcript*ome* analyses [90]. After fixation, tissues are dehydrated with alcohol/xylene (formalin) or acetone (HOPE) and the tissue is embedded in a block of paraffin to enable sections. For light-microscopic analysis of fixed tissues, a thin section, with a few micrometers thickness, is cut from the tissue. To visualize different morphological entities on the tissue sections, several staining techniques are available. One of the most frequently applied stains is hematoxylin and eosin (HE) which stains nucleic acids in blue and proteins in red/orange [91, 92]. HE staining provides high resolving information of the tissue morphology and is therefore used for histopathologic diagnostics. However, information about specific molecule classes and functional information is not obtained. For this purpose, immunohistochemistry allows selective staining of specific entities on the tissue section by incubation with antibodies. For instance, this methodology enables localization of lung surfactant proteins on tissue sections and in AECII [93, 94].

Histochemical and immunohistochemical characterization of lipids remains a challenge. It is known that sample preparation steps including the fixation, dehydration with organic solvents and the embedding of the tissues remove 73 to 91 % of the tissue lipids [95]. Moreover, during sample preparation and handling, lipids may diffuse and delocalize on the tissue which makes it necessary to use specific fixatives to immobilize the lipids [96] whereas the use of fresh frozen tissues and preparation of cryo-sections showed higher preservation of lipids within the tissues [95]. Specialized dyes are available to stain lipids such as Sudan black B, Nile blue A and Oil red O [97, 98]. Sudan black B and Nile blue A are general stains for lipids, whereas Oil red O selectively stains neutral lipids such as TAGs and cholesterols which allows to visualize lipid droplets [99]. Additionally, immunohistochemistry enables staining of specific lipid classes such as PS molecules [95]. However, to gain a more detailed view of the molecular distribution of lipids in tissue samples, these staining techniques are not specific enough since it is not possible to analyze the distribution of a single molecular species.

1.3.2 Localization of individual lipid metabolites by mass spectrometric imaging

MSI techniques are capable to monitor the local distribution of single lipid metabolites on tissue sections. The most widely applied technique is MALDI-MSI [100]. MALDI is a soft ionization technique for MS and allows desorption and ionization of intact biomolecules. For MALDI-MS, the sample is co-crystallized with an organic matrix substance which is capable to absorb energy from a laser pulse [101]. For MSI, the matrix substance is applied to cryo-sections of tissue samples which are then sequentially raster scanned with laser pulses, which enables localization of specific chemical compounds on the tissue section by their specific mass-to-charge ratio (*m/z*).

This approach was used to determine NSCLC regions on tissue slices containing tumor and tumor-free compartments [4]. Furthermore, lipids were identified that were specifically localized in tumor cells and necrotic regions of breast cancer tissues [76]. However, signal intensities depend on desorption and ionization efficiency of specific molecules and how they interact with the matrix molecules which are heavily dependent on the analyte and its environment in the tissue. Most likely, these conditions are not equal in different entities of the heterogeneous tissue morphology. Therefore, MALDI-MSI is considered not to be suitable for quantification. Furthermore, the interaction of the matrix, which is solubilized in a organic solvent, and lipid molecules might delocalize them of the tissues [96]. For these reasons MALDI-MSI is controversially discussed within the scientific community.

Taken together, these localization techniques are capable to provide valuable information about the spacial distribution of biomolecules in tissue samples. However, the localization techniques do not provide quantitative information. Therefore, target molecules are often extracted from tissues to perform sophisticated analytical procedures.

1.3.3 Lipid extraction procedures cover a broad range of lipid classes

Parts of this section were published in: Eggers and Schwudke, Lipid Extraction: Basics of the Methyl-tert-Butyl Ether Extraction, *Encyclopedia of Lipidomics*, 2016 [102].

Molecular structures and distributions of lipid molecules were analyzed in tissues which often were homogenized and lipids were extracted. A disadvantage of this procedure is that the tissue

structure is destroyed after homogenization. However, it enables to apply several chemical and physical techniques on the extracts for thorough investigation of the molecular composition of lipids.

At the beginning of the 20th century it was a big challenge to achieve pure lipid extracts from biological samples. A commonly applied method was introduced by Bloor and was a complicated and time consuming procedure based on ethanol, ether and chloroform [103]. Moreover, the extracts were not free from non-lipid contamination. Lipid analysis was revolutionized in the 1950s with the invention of the methods by Folch [8, 104] and Bligh & Dyer [9]. For the first time, these protocols enabled fast and simple preparation of lipid extracts with high purity. The extraction procedures covered a broad range of GL, GPL, SL and cholesterol species with recovery rates close to 100 % [105]. The procedures are based on monophasic solubilization of lipids from the biological material with a methanol/chloroform mixture followed by washing steps with water or saline (Fig 1.6, a and b). These methods became popular and remain routine methods for lipid analysis up to date. Therefore, the original papers of Folch and Bligh & Dyer are among the top 100 most cited reaseach publications [106].

In 2008, Matyash *et al.* [105] published an extraction protocol based on *tert*-butyl methyl ether (MTBE) as unpolar component. The principle is similar to the protocols of Folch and Bligh & Dyer but utilizing MTBE instead of chloroform. The biological material such as plasma [10] and tissue homogenate [109] is mixed with MTBE and methanol (10:3; v/v) to solubilize small metabolites and to precipitate proteins as well as other macromolecules. Afterwards, phase separation is induced by addition of water to the crude monophasic extracts (Fig. 1.6, c). The distribution of lipids between the two phases can be described by Nernst's distribution law [110]. Nernst's law states that substances distribute according to their solubility in both phases. Accordingly, this washing step removes polar molecules including amino acids and carbohydrates from the organic MTBE phase whereas lipids remain within.

Although the MTBE extraction and the Folch and Bligh & Dyer protocols are based on the same physicochemical principles, there is one important difference: The order of organic and water/methanol phase is inverted. In contrast to chloroform, MTBE and methanol have lower densities than water (MTBE: 0.74 g/mL and Methanol: 0.79 g/mL). Therefore, the lipid containing phase is located above the water phase. As consequence, the organic phase can be collected without touching the water phase or non-extractable residues such as precipitated

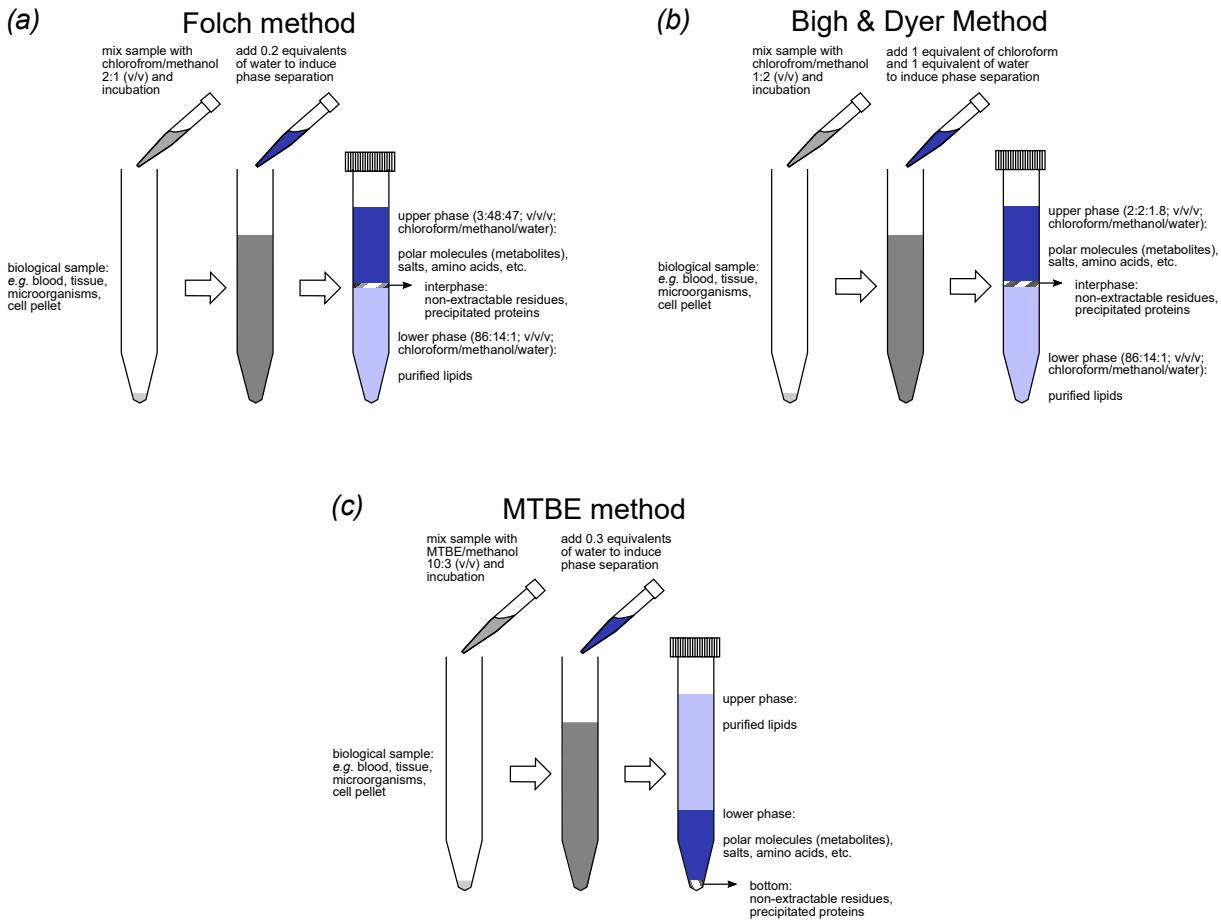


Figure 1.6: **Overview of the extraction procedures.** *(a)* Folch lipid extraction (modified from Eggers and Schwudke [107]), *(b)* Bligh & Dyer extraction (modified from Sündermann *et al.* [108]) and *(c)* MTBE extraction (modified from Eggers and Schwudke [102]).

proteins and other macromolecules (Fig. 1.6, c). This enables to perform high-throughput lipid*omics* studies with automatic pipetting systems utilizing the MTBE procedure. Another difference to chloroform is that MTBE has a higher polarity compared to chloroform and is therefore capable to take up more water (1.4 %) without phase separation [111]. The higher water content in the organic phase improves the extraction efficiency for acidic lipid classes such as PA and PG. Furthermore, MTBE is cheaper and less toxic compared to chloroform. Due to all of these advantages the MTBE protocol is widely applied in lipid*omics* and has become one of the standard procedures.

1.3.4 Electrospray ionization mass spectrometry enables quantitative lipid*omics*

Several techniques are available to analyze lipids from organic extracts. One basic technique is thin-layer chromatography (TLC) which is applied until today for rapid analysis of the lipid composition. TLC allows separation lipid classes and enables comparison the lipid class profiles in different tissues [112]. For further characterization of the lipid composition within a class, the bands on the TLC plate are scratched out and lipids are re-extracted from the obtained chromatographic material. The total FA composition within a sample can be identified by gas chromatography (GC)-MS. Sample preparation for this technique is relatively time consuming since FA-esters have to be hydrolyzed to cleave the FAs from the lipid molecules and the released FAs are further derivatized to FA methyl esters to transform them into volatile compounds that are suitable for GC-MS [113].

Until the 1980s it was not possible to analyze the chemical composition of complex lipid molecules by MS, *i.e.* to determine the FA composition of a single GPL species. This dramatically changed with development of soft ionization techniques such as MALDI and electrospray ionization (ESI). Besides MALDI, ESI is today one of the most important soft ionization techniques [114–116]. The electrospray process is conducted under atmospheric pressure and results in mild ionization conditions with little molecular fragmentation. The analyte is solved or diluted in a polar solvent which can act as proton donor. Slightly acidic conditions are advantageous to form protonated ion species, which is why acetic acid or formic acid is often added to the solution. The analyte solution is passed through a capillary with a flow-rate of 1 µL/min up to 1 mL/min, which is located in a cylindrical electrode (Fig. 1.7). An electric field of about 4,000 kV is applied between needle and mass spectrometer inlet [116]. Driven by electrostatic forces, charged droplets exit the capillary at its end where a so-called Taylor cone is formed. The ESI process is most efficient for surface active molecules such as GPLs since these molecules localize at the charged surface of the droplets. Other molecules such as sugars and unpolar lipids (*e. g.* FC) do not localize at the droplet surface and therefore, the ionization efficiency of these moelcules is orders of a magnitude lower [117]. Heated nitrogen gas is applied to evaporate the solvent from the droplets which concentrates the charges. When the charge-density at the surface comes close the maximum charge-density on a surface (Rayleigh limit [118]),

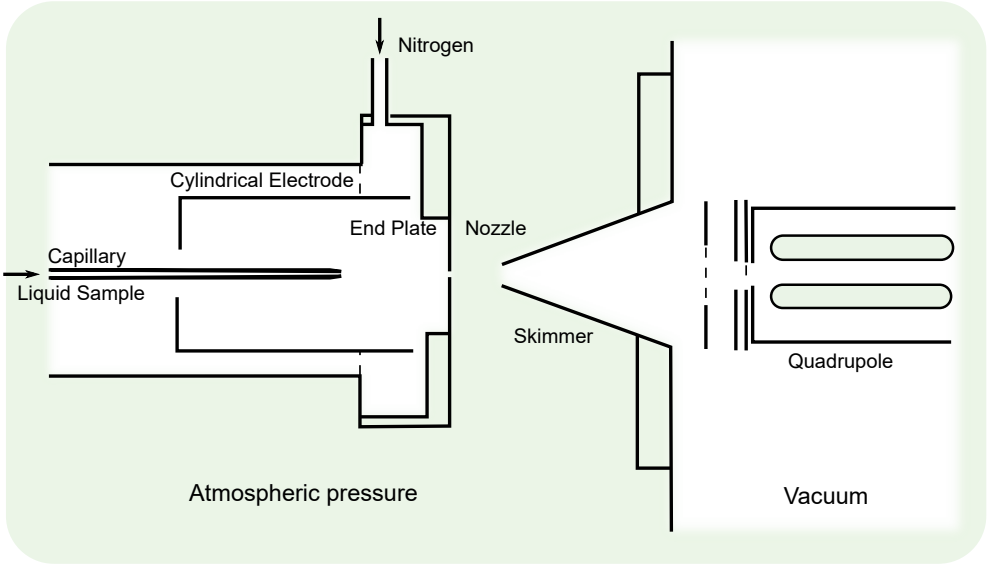


Figure 1.7: **Schematic view of an electrospray ion source.** Modified from Yamashita and Fenn [114].

smaller daughter droplets exit the mother droplets. With this process the surface-to-bulk ratio gets larger which means that a higher surface area is available. Therefore, the charges have more space to distribute on the surface which avoids reaching the Rayleigh limit. This whole process is continued until a single analyte ion remains in a droplet whose solvent evaporates to create the charged molecule in the gas phase [117]. Molecules may ionize as protonated molecular ions ($[M+H]^+$), multiple protonated species ($[M+nH]^{n+}$) or as cation adduct ions in the positive ion mode. In the negative ion mode, deprotonated molecular ions ($[M-H]^-$) and anion adducts are preferred. Due to the fact that ESI works with solutions it is suitable for quantitative analyses since ionization conditions are stable for all analytes and reference standards may be added (section 1.3.7, page 37). Furthermore, it is possible to couple this ionization technique online to separation approaches such as liquid chromatography (LC) and capillary electrophoresis. After ionization, the ions are guided through an interface into the vacuum system of the mass spectrometer where they can be analyzed by a variety of mass spectrometric techniques (Fig. 1.7).

The capillaries utilized in normal ESI ion sources have inner diameters of about 10 to 100 µm. In these applications, solvent flow-rates in the microliter per minute range are needed which easily consumes relatively large sample volumes. For applications where only a limited sample volume is available, as it is often the case when dealing with biological samples, a miniaturized

version is available: nano-ESI. Classically, for nano-ESI pulled glass capillaries are utilized which are metal (*e. g.* gold) coated at the outside and have inner diameters of 1 to 10 µm [119]. This setup enables flow rates of 20 to 40 nL/min which allows long and stable electrospray conditions while only consuming a few microliters of sample volume [119]. Therefore, even with limited sample volumes, many mass spectrometric experiments can be performed utilizing nano-ESI. These capillaries are only used for one sample which prevents sample cross contamination, carry-over and memory effects [117]. However, nano-ESI is not just a miniaturization of ESI, the down-scaling has further consequences on the ionization efficiency. The charged droplets that exit the nano-ESI capillary at the Taylor cone have diameters below 200 nm and are 100 to 1000 times smaller compared to normal ESI [119]. Due to the smaller droplets, the surface-to-bulk ratio is larger resulting that approximately all molecules have contact to the surface. The consequence is that less solvent needs to be evaporated from the droplets and the exit of the capillary can be placed closer to the mass spectrometer inlet (1-2 mm [119]) [117]. This enhances ionization efficiency and sensitivity for molecules with low surface active properties such as sugars and neutral lipids. Overall, nano-scaling of ESI heavily increases the sensitivity and also the stability against salt contamination.

However, the experimental setup of nano-ESI is relatively time consuming since each sample has to be manually loaded into the capillaries. Afterwards, the position of each single capillary tip has to be adjusted in front to the mass spectrometer inlet to achieve optimal spraying conditions. With this standard nano-ESI setup, large scale studies with hundreds of samples are hardly conducted. To speed-up nano-ESI analyses and to enhance reproducibility in regard of capillary quality and positioning in front of the mass spectromter, there were attempts to automate nano-ESI. Therefore, multiple nano-ESI nozzles were etched on a monolithic silicon wafer [120]. Such prepared wafer containing 100 to 400 individual nano-ESI nozzles is mounted to a chip and placed in front of the mass spectrometer. For each individual sample a new nozzle is supplied. This setup enables fully automated nano-ESI measurements and makes it possible to run multiple samples under stable analytical conditions. These fully automated nano-ESI robots are commercially available and routinely used in lipid*omics*.

1.3.5 Shotgun lipid*omics* as approach for analysis of lipids in biological materials

As variable as the portfolio of available mass spectrometric techniques for lipid analysis, are the approaches for lipid quantification from biological lipid extracts. Han and Gross showed that it is possible to analyze lipid*omes* from crude lipid extracts without further sample preparation [13]. This approach is known as *shotgun lipidomics* and stands in contrast to other approaches utilizing separation techniques coupled to ESI-MS such as LC-MS. LC-MS has the advantage that ion suppression effects are lowered since contaminations and ion suppressing agents are separated from the analytes of interest which increases the dynamic range for quantification. Disadvantages of LC-MS based lipid*omics* are the relatively time-consuming analytical procedures. Furthermore, lipids strongly adhere to chromatographic materials and therefore cannot be completely washed from the chromatographic column during an analytical run. This behavior could cause sample carry-over and memory effects, which are often not thoroughly examined.

In shotgun lipid*omics*, lipid extracts are directly infused into the mass spectrometer without any prior separation of the crude lipid extract. Automated sample infusion techniques are available for instance by automated flow injection [121] where the sample is injected into an isocratic solvent flow or else by automated chip-based nano-ESI [12, 14, 89, 122]. The advantage of shotgun lipid*omics* is, that it is fast and covers a broad range of GLs, GPL, SL and cholesterols. Usually, acquisitions are performed in the positive *and* in the negative ion mode since not all lipid classes are accessible from a single ion mode. Shotgun lipid*omics* data acquisitions in both ion modes are finished in less than 10 minutes with modern instrumentations [122], and quantification results are stable. However, even in shotgun lipid*omics* different approaches are used for lipid analysis utilizing different instrumentations. The basic approaches of shotgun lipid*omics* are *bottom-up*, which is based on lipid identification by tandem mass spectrometry (MS^2), and *top-down*, which identifies lipid species by accurate mass determination.

***Bottom-up* shotgun lipid*omics* identifies lipid species by specific fragmentation patterns.** The basic *bottom-up* shotgun lipid*omics* experiments were performed on a triple quadruple mass spectrometer (QqQ). The QqQ combines two quadrupole (Q) mass filters with a colli-

sion cell (q) to perform MS^2 experiments. The first quadrupole (Q_1) acts as mass filter and can either be scanning a mass range or be fixed to a specific *m/z*. The second quadrupole (q) can be filled with a neutral gas for collision-induced dissociation (CID) to fragment the selected ion species. The third quadruple (Q_3), again, is a mass filter that can either be scanning or fixed. Depending on the lipid class, specific fragment ions or neutral losses from precursor ions can be generated by CID. For example, lipids with choline HGs such as PCs and SMs generate a characteristic fragment of *m/z* 184 in the positive ion mode referring to the choline HG. For precursor ion scan (PIS), Q_1 is scanning, Q_3 is fixed at *m/z* 184, and a mass spectrum is drawn of the ion intensity of *m/z* 184 over the Q_1 mass. The resulting mass spectrum exclusively contains signals from PCs and SMs. Similarly, PI is identified by a specific fragment ion at *m/z* 241 in the negative ion mode [123]. When the lipid HG does not provide specific signals, but loses a characteristic structural entity, neutral-loss scanning (NLS) can be applied. For example, PE molecules have a neutral loss of 141 Da in positive ion mode which represents the loss of the phosphoethanolamine HG. This means that PE molecules lose 141 Da in their molecular weight after passing the CID cell (q). In this example of NLS, Q_1 and Q_3, both are scanning, with an offset of 141 Da. Afterwards, a mass spectrum is generated plotting the ion abundance over the Q_1 center mass and the achieved spectrum shows signals from PE species. Similarly, PS is identified with a neutral loss of 185 Da in the positive ion mode. In this manner, a comprehensive picture of the lipid*ome* is drawn by sequential PIS and NLS experiments [123]. This approach is widely applied but has some limitations. Due to the low selectivity of the Q mass analyzer of about 1 Da selection window, PC and SM are not distinguished in the positive ion mode. Lipids with *O*-acyl and *O*-alkyl (ether lipids) linked aliphatic chains cannot be separately detected. The third limitation is that not all lipid classes provide characteristic signals or neutral losses such as PA and PG.

***Top-down* approach is based on high-resolution mass spectrometry.** Lipid identification is based on accurate mass determination which enables to determine the chemical sum composition of each ion species from its accurately determined *m/z*. For *top-down* lipid*omics* screens, the mass resolution of the utilized instrument needs to be high enough to separate lipid species with same nominal masses (isobaric), but with different exact masses. An example for two isobaric lipid species is PC 34:1 (*m/z* 760.585082 $[M+H]^+$) and PC *O*- 35:1

(*m/z* 760.621467 $[M+H]^+$). Formally, in PC *O*- 35:1 an oxygen atom (15.994366 atomic mass units (amu)) is exchanged with CH_4 (16.030752 amu) which results in a mass difference of 0.036386 amu between both species. An instrument suitable for *top-down* lipid*omics* screens has to resolve at least this mass difference. The resolving power (RP) of a mass spectrometer is defined as the signals' *m/z* divided by a mass difference (Δm). Δm is the full-width-at-half-maximum (FWHM) of the respective signal. To resolve PC 34:1 and PC *O*- 35:1, the FWHM of the signals has to be smaller than the mass difference of both signals (0.036386 amu). Consequently, an RP of at least 21,000 is needed in *top-down* lipid*omics*. An RP of at least 42,000 is necessary for baseline separation. Most utilized instruments have an RP of 100,000 at *m/z* 400 or even higher. Mass analyzers that provide high RP are Orbitrap [124], Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometers [125] and modern time of flight (TOF) instruments [126].

High mass resolution together with high mass accuracy enables relative errors for mass determination lower than 2 parts-per-million (ppm) which is needed to calculate the chemical sum composition of the ion species [14, 30]. Starting from the specified chemical sum composition of a signal with accurately determined *m/z*, lipid species are assigned. The lipid class is determined from the number of heteroatoms in a chemical formula. For example, PC $[M+H]^+$ has a heteroatom sum composition of $O_8N_1P_1$ which is distinguishable from SM $[M+H]^+$ with the sum composition $O_6N_2P_1$ [30]. Furthermore, the numbers of carbons and hydrogens are known within the HGs, which are $C_{10}H_{20}$ for PC and $C_{11}H_{23}$ for SM. After subtracting the number of carbon and hydrogen atoms present in the HG from the determined molecular sum, the numbers of carbons and hydrogens within the aliphatic chains remain, which allows identification of the sum composition of the FA residues. The chain-length is determined by the number carbon atoms. The number of dbs is calculated by the ratio of carbons and hydrogens (double-bonds and ring, dbr). With this methodology, a broad range of GLs, GPLs, SLs and cholesterols can be identified from a single high resolution survey mass spectrum. However, there are still cases where heteroatom sum compositions of lipid classes overlap such as for PC $[M+H]^+$, PE $[M+H]^+$ and PA ammonium cation adduct ions ($[M+NH_4]^+$) which all have $O_8N_1P_1$ in the positive ion mode. Therefore, it has to be carefully verified which ion species and which ion mode is suitable for quantification *top-down* screens to avoid signal interference.

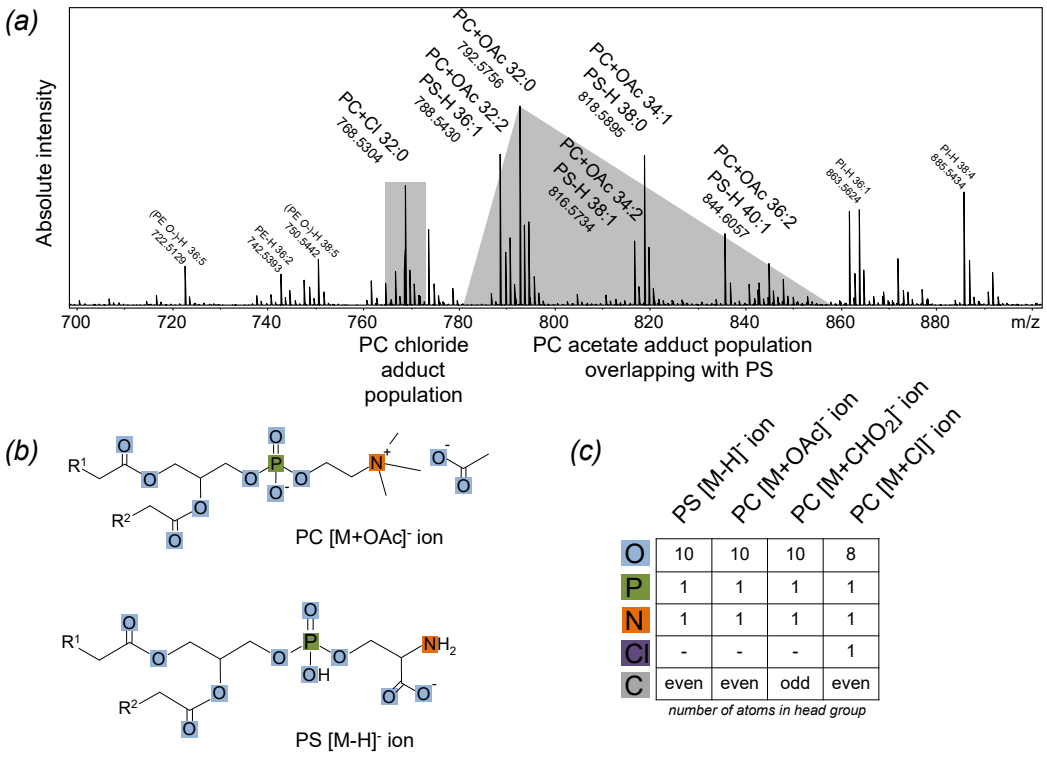


Figure 1.8: **Overlapping signals using ammonium acetate.** *(a)* Survey mass spectrum shows overlapping of PC $[M+OAc]^-$ with PS $[M-H]^-$. Additionally, splitting into two ion populations of PC is shown: $[M+OAc]^-$ and $[M+Cl]^-$ ions. *(b)* Chemical structures of PC and PS ions which have the same heteroatom sum composition. *(c)* Heteroatom sum compositions of PS $[M-H]^-$, PC $[M+OAc]^-$, $[M+CHO_2]^-$ and PC $[M+Cl]^-$.

1.3.6 Salt additives can be used as tool to adjust the ionization conditions

The additive ammonium acetate (NH_4OAc) is commonly used to support ionization processes in shotgun lipid*omics* experiments. The presence of ammonium (NH_4^+) cations in the sample matrix is advantageous since neutral lipids such as TAG, DAG, and CE, form $[M+NH_4]^+$ ions [30]. This allows mass spectrometric detection of species that hardly ionize as $[M+H]^+$ ions and increases sensitivity for them. In the negative ion mode, acetate adduct ions ($[M+OAc]^-$) out of SM, PC, Cer as well as out of the neutral lipids DAG and TAG are formed and enable detection of these species there although the HGs are preferably positively charged. However, PC $[M+OAc]^-$ ions overlap with PS $[M-H]^-$ since actetate (OAc^-) is added to an ion species which adds two oxygen atoms to the overall sum composition of the respective lipid. Specifically, this changes the heteroatom sum composition of a PC $[M+OAc]^-$ ion from O_8PN to $O_{10}PN$ (Fig.

1.8, b and c) which is the same heteroatom sum composition as expected for PS $[M\text{-}H]^-$ (Fig. 1.8, b and c). Therefore, both PS and PC have overlapping mass spectrometric signals which are not distinguishable by high resolution MS (Fig. 1.8, a).

NH_4^+ formate (CHO_2^-) and chloride are potential alternative salt additives (Fig. 1.8, c). The addition of CHO_2^- instead of OAc^- does not change the heteroatom sum composition but the number of carbon atoms in the PC-HG changes from an even number to an odd number for PC formate adduct ions ($[M+CHO_2]^-$) (Fig. 1.8, c). Under the assumption that exclusively even numbered FAs are present, unambiguous identification of PC $[M+CHO_2]^-$ and PS $[M\text{-}H]^-$ is possible [122]. In this manner, the chemical sum compositions of PS' are expected with even numbers of carbon atoms (Fig. 1.8, c) and PC molecules are expected with overall odd numbers of carbon atoms. However, it was already introduced that low abundances of odd numbered FAs are present even in mammalian systems (section 1.2.1, page 15). The key to solve this problem is, to force PC to form adducts with an anion changing the heteroatom sum composition of the entire ion population. A promising candidate is ammonium chloride (NH_4Cl) which changes the heteroatom sum composition of PC and yields in characteristic signals for both, PC and PS in the negative ion mode (Fig. 1.8, c).

1.3.7 Lipid quantification

After assignment of the mass spectrometric signals to specific lipid species by their accurately determined *m/z*, lipid quantities are calculated based on the signal intensities. To refer from the absolute signal intensities to a molar quantity of the respective lipid species, each signal is referenced to characteristic peaks from internal standards (ISDs). ISDs are reference substances added to the samples at the earliest possible step in sample processing [105, 127]. ISDs have to be detected unambiguously by the analytical system in samples containing complex matrices and their chemical and physical properties need to be as close as possible to the molecules of interest. This procedure covers potential losses of analytes, for example during extraction processes. In mass spectrometric analyses, utilization of ISDs is required to cover ion suppression effects during the ionization process by signal normalization to the ISD intensities.

A well suited ISD solution would comprise isotope labeled lipids from each class for instance with deuterium, ^2H (D). Hydrogen and D have a mass difference of 1.0063 amu which makes

D-labeled standard substances unambiguously detectable by their accurate mass. Practically, a complete set of isotope labeled standard substances is often not achievable due to their high cost. Consecutively, a different approach for standard selection can be used. Lipid substances with either odd numbered FAs such as FA 17:0 or lipids with two *O*-alkyl linked side chains (diether lipids) can be utilized as ISDs in human. These types of lipids are not available by mammalian biosynthetic pathways and therefore considered to be not present in human samples.

Based on the signal intensities of ISDs and their known concentrations, the analyte concentration is estimated by putting the respective ion intensities of analytes and standards into a relation. However, it has to be considered that this strategy is an approximation. In practice, all species from a lipid class are often referred to a single standard substance from the same class since lipids from the same class undergo the same ionization mechanisms [30, 127].

1.4 Analysis strategies for lipid*omics* data

Various strategies have been applied for the analysis of lipid*omics* data. This section provides a brief overview of techniques utilized in lipid*omics* research. Hypothesis testing by the t-test or the Wilcoxon Mann-Whitney U (U)-test is introduced as method to compare lipid*omes* between groups. Unbiased multivariate statistical methods such as principal component analysis (*pca*) and hierarchical cluster analysis are suitable to evaluate relationships and similarities within lipid profiles. Linear partial least-squares (PLS) regression is presented as a tool for the association of lipid*omics* data with the tissue phenotype and clinical parameters. Finally, this section introduces the concept of lipid*ome* homology.

1.4.1 Statistical tests identify significantly altered lipid species

Statistical tests such as the *t*-test and the U test are widely used to investigate specific lipid*ome* alterations [128–131]. Which test has to be chosen depends on the data structure. The t-test is a good choice for normal distributed data [132], especially when the variances are similar in the compared groups. Non-parametric tests as the U test do not assume normal distribution but put the data into ranks [133] which makes them applicable without pre-assumptions regarding data structure and are therefore a good choice for not normal distributed data and for data with

highly differing variances. Practically, it has to be considered that the U test needs at least four to five observations per group. Otherwise, no significant results are returned from the test. For each variable one U test is calculated comparing two levels, *i. e.* two groups compared by phenotype (treatment *vs.* control). Lipid*omics* data contain, in a similar manner to other *omics* disciplines, some hundreds of lipid species (variables) in often less than one hundred observations (samples). Therefore, it is necessary to perform multiple tests and the level of significance has to be adjusted in order to avoid miss-interpretation of the results. Methods for such *p*-value adjustments are for example the Bonferroni adjustment [134] and the Benjamini-Hochberg false discovery rate (FDR) [135]. Both, multiple t-tests and U tests can be visualized in a volcano plot. A volcano plot presents the fold change of each variable between two levels and is plotted over the *p*-value from the respective test. This approach can for example be applied for biomarker discovery [136]. The t-test and the U test are limited to compare two populations, *e. g.* disease and healthy control. In case that the dataset is more complex and there are more levels to compare, other methods for such comparisons are available such as analysis of variance (ANOVA) which assumes normal distributions [132] and the non-parametric Kruskal-Wallis test [137].

1.4.2 Correlation based analyses allow hypothesis-free and multivariate data interpretation

Unsupervised multivariate statistical tools analyze large datasets without further information regarding grouping or classification. Unsupervised methods are capable to analyze natural similarities and homologies and are frequently applied for evaluation of *omics* data [138]. These methods include *pca* [30] that transforms a multivariate dataset into a few latent variables expressing the highest level of variability within the dataset. Hierarchical clustering enables calculation of correlations between the individual lipid*omes* visualized by a hierarchical tree (dendrogram) as it is often performed to analyze gen*omics* data [138]. Unsupervised methods provide the possibility to analyze the data for instance to test whether similar phenotypes form closely related clusters.

Other methods to test for specific lipid*ome* alterations between predefined groups are supervised correlation based techniques such as principal component discriminant analysis and

PLS-discriminant analysis (DA) [4, 139]. These methods are suitable to identify lipid*ome* alterations along a categorical discriminator variable by transforming multiple variables into a few latent variables known as principal components (*pc*s). Moreover, the *pc*s are correlated to a discriminator variable. For instance, PLS-DA was used to discriminate lipid*omes* of obese mice from wild type mice and further to separate males from females. Furthermore, specific alterations in the TAG levels were discovered in obese mice [140].

1.4.3 Regression analyses enable association of lipid*omes* with clinical parameters

Regression analyses enable to reveal relationships between lipid*omes* and continuous variables such as the BMI and the age of the respective subject. PLS regression is an approach to calculate multiple regressions based on a PLS algorithm. The methodology was introduced by Bruce Kowalski and Svante Wold in the field of *chemometrics* to analyze complex spectroscopic data [141, 142] but has also been widely applied to analyze *omics* data since then [143]. With this approach, response variables are explained by predictor variables in a linear regression model while transforming the original variables into a new set of latent variables: components calculated for predictor variables (t). The t-components are calculated to express the variability within the dataset and further to correlate them to the responses. Overall, PLS regression is an approach to generate stable multivariate linear regression models. However, there is still the need to validate the outcomes of these regression models by cross-validation (cv) approaches. There, the data is split in a training set for model generation a validation set to evaluate the model on an independent dataset. This helps to prevent the computed model from overfitting, verify its robustness and avoid misinterpretations.

1.4.4 Lipid*ome* homology is a powerful concept to investigate structural alterations

All the statistical methods listed before are dependent on quantitative lipid*ome* data. In this regard, differences in lipid quantities are associated to specific disease states or a specific phenotype. However, only lipid species can be compared that are present in each phenotype. If

parameters are missing, none of these methodologies can be applied. There are approaches to impute missing values, starting with the easiest approach to impute overall mean values, to more sophisticated strategies using iterative algorithms [144]. However, imputation of missing values can only be performed for variables with valid quantitative values in more than 80 % of all samples. Otherwise, the imputation error would falsify the results. Lipids that are present in less than 80 % of the observation cannot be further analyzed. If general reprogramming is assumed, it could be possible that specific metabolic pathways are switched on or off, yielding in the observation that some lipids only occur in a specific phenotype. This information would not be available after filtering. To overcome these limitations, Marella *et al.* proposed a way to analyze lipid*omes* based on their structural similarity and not on lipid quantities [145]. Therefore, lipid structures are translated into a chemical space model for each lipid*ome*. Afterwards, the similarity between two lipid*omes* is expressed by the lipidome juxtaposition (LUX) score as a measure for lipid*ome* homology [145]. Utilizing this methodology, it was shown that it is possible to visualize specific lipid metabolic alterations in yeast and Drosophila models [145].

1.5 Objectives

With this dissertation lipid metabolic processes of human lung tissues are investigated utilizing tissue samples from lung cancer patients after surgery. This study aimed to identify specific features in the lipid*omes* that changed between tumor-free alveolar lung tissues corresponding tumors. It was investigated whether the lipid profiles of the main entities of NSCLC, ADC and SCC, were distinct and whether lipid*omes* of NSCLC and carcinoid tumors differed. An important part of this study deals with the association of the lipid*ome* data with histopathological characterizations of the tissue phenotype and with clinical parameters.

Clinical parameters such as age, gender, BMI, smoking behavior as well as quantitative scores for the histological phenotype were assessed in this study. With this information, it was aimed to systematically evaluate which multivariate statistical methods are applicable to relate lipid*omics* data with individual clinical and histological characteristics.

The largest alterations were expected to be localized between tumor lipid*omes* and the corresponding tumor-free alveolar tissues. Furthermore, the histologic phenotype is expected to induce observable lipid*ome* alterations whereas the influence of the patients' age, smoking, gender and other lifestyle parameters might be much more ambiguous. For this study, several available data analysis strategies were tested to follow the objective to evaluate and discover possible lipid*ome* influencing parameters starting from the main impacts to lightly nuanced alterations in the context of lifestyle parameters.

2 Results

This study aimed a achieve a most comprehensive snapshot of the human lung lipid*ome*. To enhance the coverage of anionic lipid species in the negative ion mode, the lipid*omics* screen approach was customized. This optimization together with the analytical characteristics of the analytical workflow are presented in the first part of this chapter. The analytical setup was then utilized in a pilot study comprising 43 tissue samples from 26 patient to investigate lipid*ome* alterations in carcinogenesis. In a follow-up study, which included 174 tissue lipid*omes* from 92 patients, the outcomes of the pilot study were further specified. This part has a strong focus on regression modeling of lipid*omics* data to evaluate potential lipid*ome* influencing parameters.

2.1 Sensitive and reproducible lipid*omics* screens

This study utilized a stable, precise and reproducible lipid quantification strategy based on lipid*omics* screens. To prevent signal interference of PC and PS in the negative ion mode and to gain sensitivity for the anionic lipid species, the salt additive for negative ESI was adjusted. NH_4Cl was selected as optimal additive and the optimization of the added salt concentration is presented in the following section. After optimization, the technical variability of the entire screening workflow was evaluated.

2.1.1 Ammonium chloride as ESI additive in the negative ion mode

The commonly used NH_4OAc additive for ESI caused an overlap of isobaric peaks in the negative ion mode, specifically for PC $[M+OAc]^-$ and PS $[M-H]^-$ (section 1.3.6, page 36). Experiments carried out with NH_4OAc as additive revealed distribution of PCs into two ion populations. In addition to an ion population of PC $[M+OAc]^-$, a second ion population of chlorinated

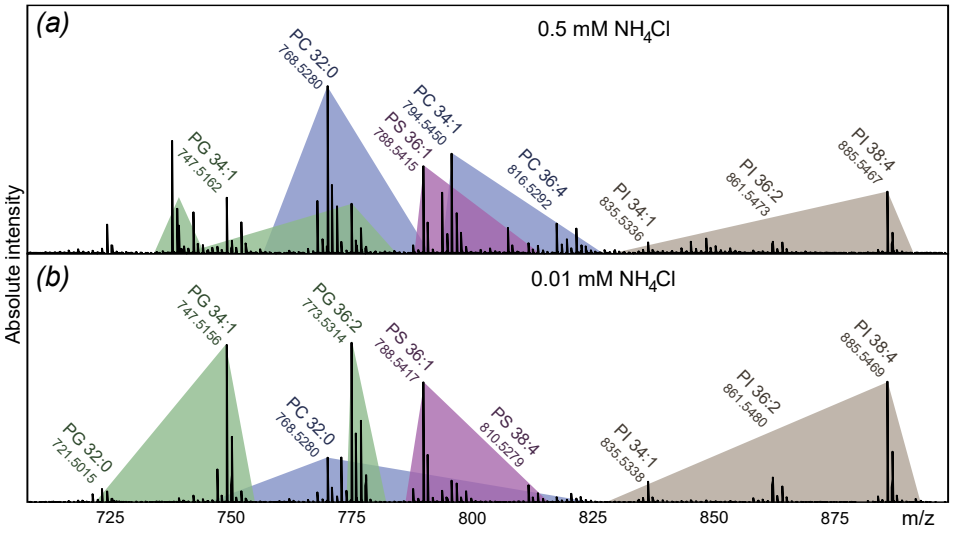


Figure 2.1: **Mass spectrometric signal profiles drastically changed with concentration of the electrospray ionization additive ammonium chloride.** *(a)* FT-ICR mass spectrum in negative ion mode using 0.5 mM NH_4Cl as ESI additive. Signals from PC are illustrated in blue, PG in green, PS in purple, and PI in brown. *(b)* Negative ion mode spectrum with 0.01 mM NH4Cl from the same sample. Modified from Eggers *et al.* [146].

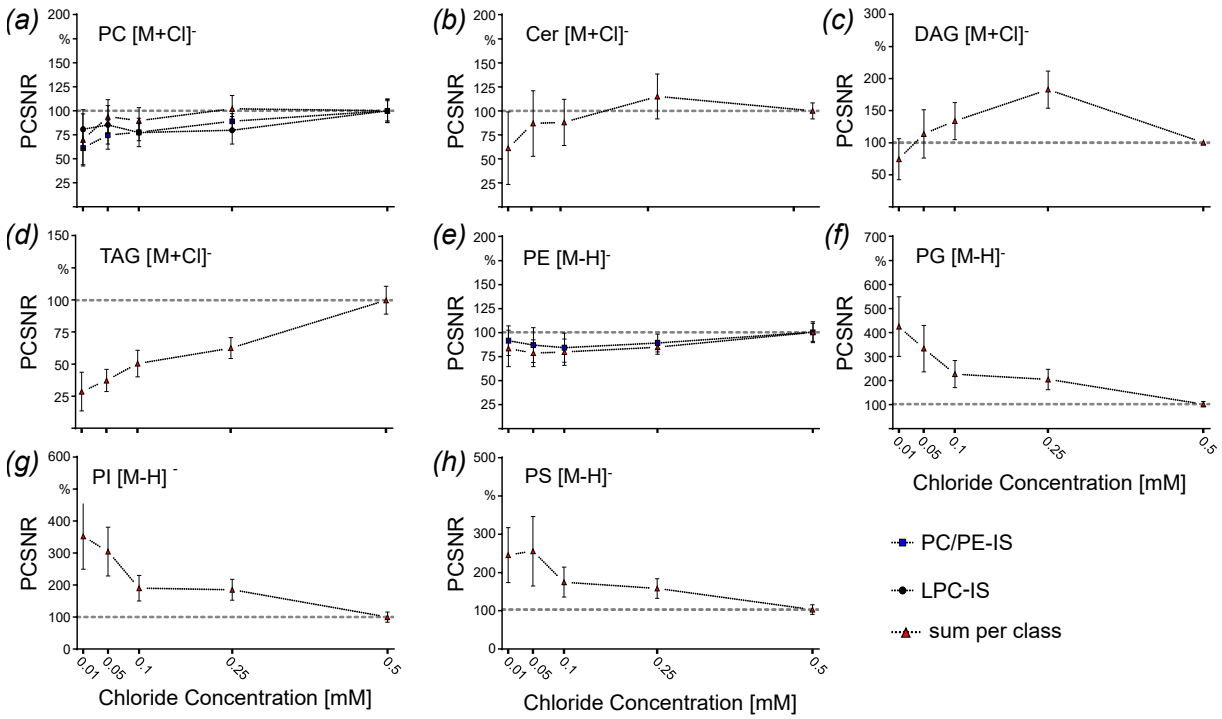


Figure 2.2: **The ammonium chloride concentration heavily influenced the sensitivity for detection.** The percentage change of signal-to-noise ratios (PCSNR) as function of the NH_4Cl conctration. The PCSNR was calculated as sum of the SNRs of the individual lipids for a given class. These values were normalized to their average response at the additive concentration of 0.5 mM NH_4Cl. Lipid classes: *(a)* PC (19 species), *(b)* Cer (12 species), *(c)* DAG (9 species), *(d)* TAG (3 species), *(e)* PE (9 species), *(f)* PG (17 species), *(g)* PI (11 species), and *(h)* PS (10 species) including SNRs for ISDs of PC, LPC and PE. Modified from Eggers *et al.* [146].

adduct ions ($[M+Cl]^-$) was observed although no chloride was added to the samples (Fig. 1.8, a, page 36). Most likely, the $[M+Cl]^-$ ion population was observed due to unavoidable carry-over of chloride from the biological matrix. By addition of NH_4Cl instead of NH_4OAc, this distribution into different ion populations was avoided. Under this condition all PC molecules formed $[M+Cl]^-$, thereby reducing spectral complexity and enhancing selectivity for determination of all PC and PS species. Due to these advantages, NH_4Cl was selected as additive for negative ionization to yield unambiguous quantification of all lipids. With addition of chloride, isobaric PC and PS species such as PC 34:1 ($[M+Cl]^-$, *m/z* 794.5472) and PS 37:5 ($[M-H]^-$, *m/z* 794.4978) had a mass difference of 0.0494 amu which is easily discriminated by high resolution MS.

First experiments utilizing NH_4Cl as additive were carried out with similar additive concentrations as routinely used in the laboratory with NH_4OAc (about 3.7 mM). Compared to the spectra obtained with NH_4OAc, spectra with NH_4Cl showed lower and less stable signal intensities. The inference from this observation was that the concentration of NH_4Cl was potentially not optimal suited for lipid*omics* screens and should be optimized.

Therefore, the NH_4Cl concentration in the spray matrix was optimized to improve stability and sensitivity. The data showed a significant shift in the ratios of signal intensities of PC and PG in the mass spectra obtained with 0.5 mM (Fig. 2.1, a) and with 0.01 mM (Fig. 2.1, b) NH_4Cl additive. In the spectrum obtained with 0.5 mM NH_4Cl, PC 32:0 had the strongest signal. However, in the spectrum with 0.01 mM NH_4Cl, PC 32:0 had only about 30 % of the intensity of the highest signal (PG 36:2). These observations showed that it was possible to adjust relative signal intensities and customize the analytical conditions with the ESI additive.

Consecutively, it was examined whether the strong changes in the raw mass spectra were also reflected in altered sensitivities for individual signals. The signal-to-noise ratios (SNRs) were used as a measure for the sensitivity (Fig. 2.2). The overall sensitivity per lipid class was expressed as percentage change of signal-to-noise ratios (PCSNR). The optimization was focused on the anionic lipid classes including PIs, PGs and PS' as well as on PCs and Cers. The neutral lipids, TAG and DAG, are easily accessible in the positive ion mode and were therefore not considered for optimization. PCSNR values systematically decreased for $[M+Cl]^-$ ion forming lipid classes with decreasing NH_4Cl additive concentrations, most clearly observed for TAGs (Fig. 2.2, d) but also for PCs (Fig 2.2, a) and Cers (Fig. 2.2, b). Sensitivities

of DAG increased from 0.5 mM to 0.25 mM NH_4Cl additive concentration, followed by a steady decline with lower NH_4Cl concentrations (Fig. 2.2, c). In contrast, other lipid classes which ionized as $[M\text{-}H]^-$ ions had a different behavior as a function of the NH_4Cl additive concentration. For instance, PEs showed approximately constant sensitivities over the entire NH_4Cl concentration range from 0.01 mM to 0.5 mM (Fig. 2.2, e). In contrast, PGs, PIs, and PS' demonstrated increased sensitivities with reduction of the NH_4Cl additive concentration (Fig. 2.2, f-h). Finally, an NH_4Cl concentration of 0.05 mM provided the highest sensitivities for PC, Cer, PG, PI and PS in optimal screening conditions.

2.1.2 Analytical characteristics of the developed workflow

The parameters of the analytical lipid quantification with regard of the technical variability of the entire workflow were tested for their stability. It was asked whether the analytical approach was sensitive enough to answer questions regarding the biological problems. The technical variability between repeated mass spectrometric acquisitions of same tissue extracts was examined. Therefore, relative standard deviations (RSDs) of repeated measurements were inspected. A list containing all measured RSDs was created which is presented in the frequency distribution (Fig. 2.3, a). The RSDs measured had values up to 140 % whereas 95 % of all RSD values were lower than 43 %. The proportion of 68 % of the measured RSD values indicated a variability of less than 13 % (Fig. 2.3, a).

Lipid quantification of repeated extractions of the same homogenates showed reproducible results (Fig. 2.3, b). Quantitative values from two independent sample preparations correlated perfectly with linear regression coefficients (R^2s) of about 0.98 or even higher. Lipid quantification was reproducible in more than 3 orders of magnitude (Fig. 2.3, b). Quantitative values larger than 100 pmol/mg tissue yielded in best correlations between repeated analyses. Lower concentration ranges showed higher variability, probably since these values were close to the detection limit.

The question was whether the analytical precision, estimated with 13 %, was sufficient to assign lipid*ome* alterations to the biological phenotype. For this purpose, it was tested whether the technical variability was lower than the variability between different samples. Therefore, a set of samples was repeatedly extracted and the data were analyzed by *pca*. The technical

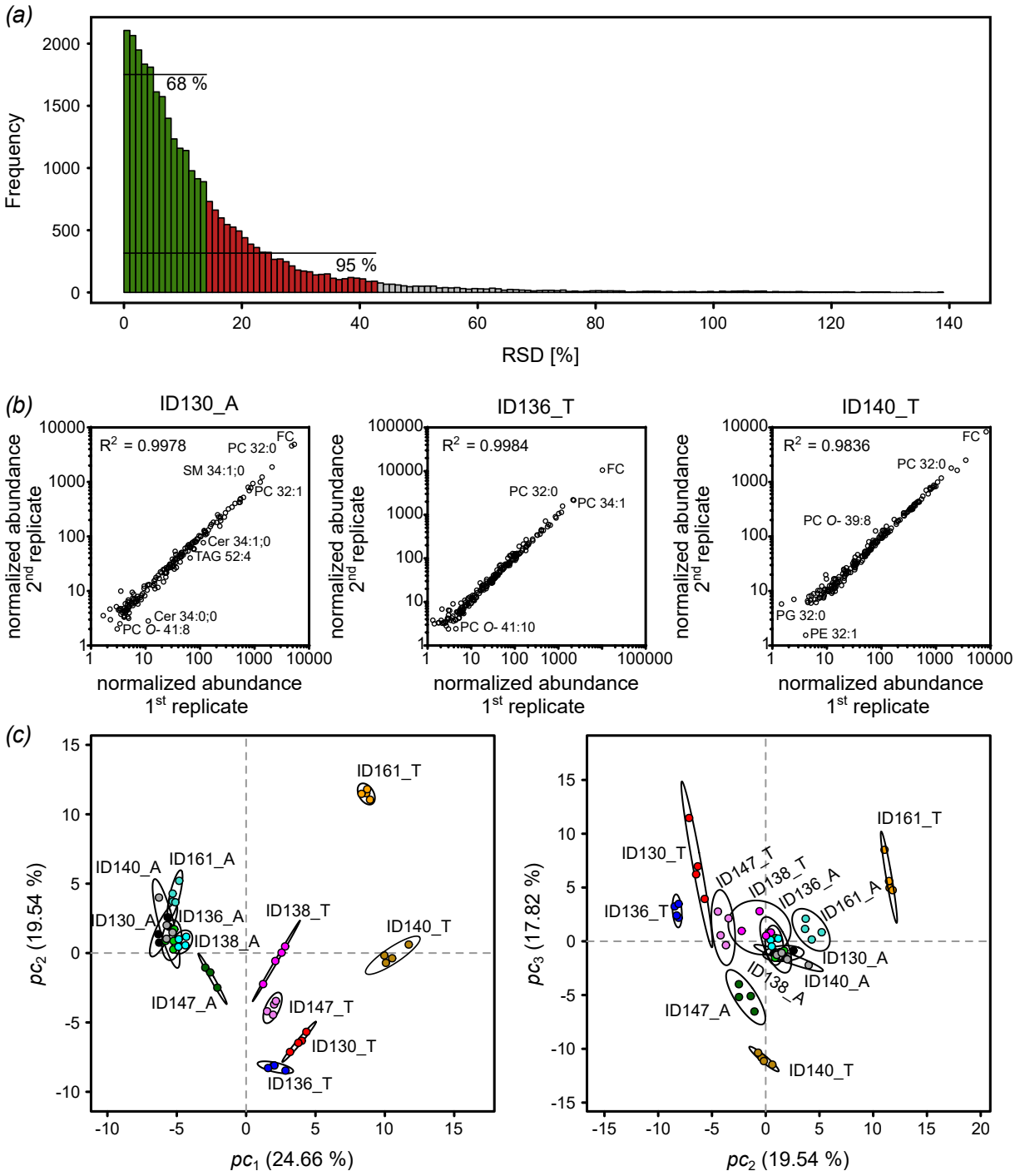


Figure 2.3: **Reproducibility of the lipid*omics* screens.** *(a)* Frequency distribution of RSDs obtained by repeated mass spectrometric acquisitions. The list contained all RSD values from the analyzed 174 tissue extracts where repeated mass spectrometric acquisitions were averaged. *(b)* Linear correlations of lipid*ome* data from two independent technical replicates from each, ID130_A, ID136_T, and ID140_T. The complete extraction procedure was repeated starting at the tissue homogenates. Lipid*omes* of both independent extractions are compared in linear regressions. R^2 values are noted in the panels. Modified from Eggers et al. [146]. *(c) pca* summarizing technical variability of the screening assay: pc_1 *vs.* pc_2 and pc_2 *vs.* pc_3. Dots in the same color are derived from the same tissue homogenate. Four dots per sample include two independent extraction procedures with two repeated mass spectrometric acquisitions. Ellipses indicate 95 % confidence areas.

variability of repeated extractions and mass spectrometric analyses appeared to be lower than the *inter*-sample variability as indicated by the confidence areas (Fig. 2.3, c).

2.2 Lipid*omes* of human lung tissues revealed distinct alterations in cancer and emphysema

At first, the developed method was applied in a pilot study. The pilot study comprised lipid*omes* of 43 tissue samples obtained from 26 patients as summarized in table 4.3 (page 102) (*sub-cohort 1*). The question was whether the lipid profiles of alveolar lung tissues and matched tumor tissues had marked differences and in which manner the histopathological tissue phenotype is reflected within the lipid*omes*. The results presented in this chapter were published in Eggers *et al*., Scientific Reports, *2017* [146].

2.2.1 Lipid*omes* of alveolar and tumor tissues had distinct molecular patterns

The first question was whether the lipid*ome* of lung tumors and tumor-free alveolar lung tissues reflected the distinct morphological features of these two tissue types (Fig. 2.4, a and b). The patient with the identifier (ID) 64 was chosen as an example to specify the alterations that were identified between alveolar and tumor tissues.

The raw mass spectra clearly showed different patterns in the signal distributions of the alveolar and the tumor tissue (Fig. 2.4, c-f). These mass spectrometric profiles alone were sufficient to discriminate alveolar and tumor tissues. The positive ion mode spectrum of the alveolar tissue extract (Fig. 2.4, c) revealed PC 32:0 (*m/z* 732.5676, $[M+H]^+$) as most intensive signal. In contrast, PC 34:1 (*m/z* 760.5858, $[M+H]^+$) was the most intense ion in the corresponding ADC-tumor tissue (Fig. 2.4, e). Unsaturated lipids such as PC 36:4 (*m/z* 782.5692, $[M+H]^+$), CE 18:2 (*m/z* 666.6187, $[M+NH_4]^+$) and TAG 52:2 (*m/z* 876.8025, $[M+NH_4]^+$) had larger signal intensities in the tumor tissue. Complementary information was obtained from the negative ion mode, since lipids with anionic charged HGs were predominantly visible there (Fig. 2.4, d and f). The alveolar tissue was characterized by intense signals of PG species such as PG 34:1 (*m/z* 747.5162, $[M-H]^-$) and PG 36:2 (*m/z* 773.5307, $[M-H]^-$) (Fig. 2.4 d). In the spectrum of

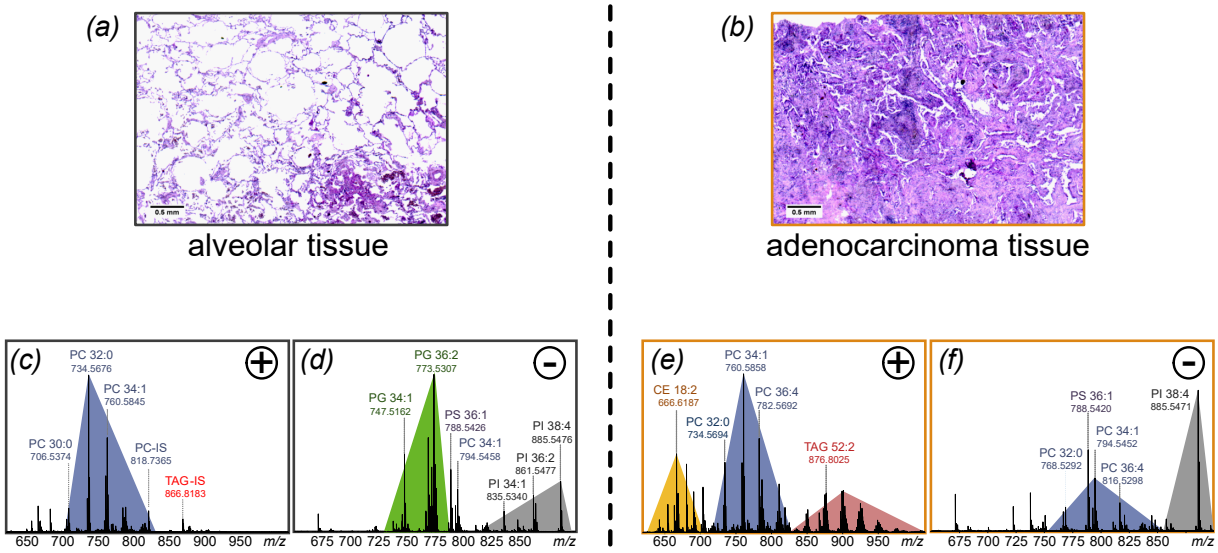


Figure 2.4: **Distinct lipid profiles in alveolar and tumor tissues.** *(a)* HE stained tissue sections of the alveolar tissue obtained from patient ID64. *(b)* Matched ADC-tumor tissue section from patient ID64. *(c)* Positive ion mode FT-ICR mass spectrum from patient ID64, alveolar tissue. *(d)* Corresponding negative ion mode spectrum of ID64, alveolar. *(e and f)* Mass spectra (positive and negative ion mode) from the ADC-tumor tissue of ID64. Mass spectra were baseline corrected and smoothed. Signals of most abundant lipid identifications are annotated and series of signals from main classes are highlighted by colored rectangles (blue: PC, green: PG, red: TAG, orange: CE, and gray: PI). Modified from Eggers *et al.* [146].

the corresponding ADC-tumor tissue, these signals were scarcely present, whereas an intense signal of PI 38:4 (*m/z* 885.5471, $[M\text{-}H]^-$) was visible (Fig. 2.4, f). These were the first indications of extensive lipid metabolic alterations between lung tumors and their corresponding alveolar tissues even on raw mass spectrum level. These observations, made with the example of patient ID64, were confirmed by a larger number of paired alveolar and tumor samples. It emerged that there were numerous lipids either present in alveolar or in tumor tissues. Therfore, lipid*omes* were analyzed utilizing the LUX score appoach by Marella *et al.* [145] which compares all chemical structures and is sensitive for lipid species present only in one of the compared phenotypes.

Using the patient ID64 as an example, the lipid*ome* alterations identified on the raw mass spectra were further specified on the lipid*ome* maps which visualize the chemical space model. The lipid*ome* of the alveolar tissue (Fig. 2.5, a) had 13 PG species that were not present in the corresponding ADC-tumor lipid*ome* (Fig. 2.5, b) whereas the ADC-tumor lipid*ome* had no unique PG species compared to the alveolar tissue. There were numerous neutral lipids from the classes TAG and CE, which were identified exclusively in the ADC-tumor lipid*ome*. In this

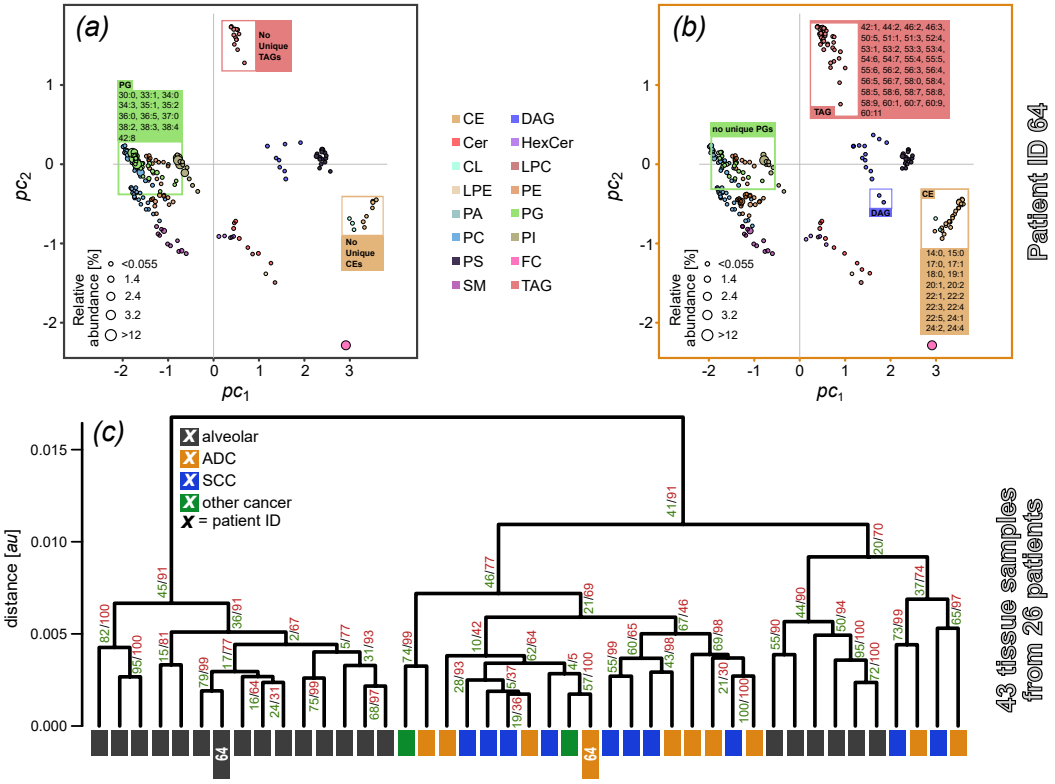


Figure 2.5: **Distinct molecular patterns in the lipid*omes* of lung tissues.** *(a)* Lipid*ome* map of patient ID64 alveolar tissue. Each point on the lipid*ome* map represents a lipid and distances between lipids represent their structural similarity. The size of the data point corresponds to the quantity (relative abundance). Uniquely identified lipids (either in tumor or alveolar tissue) of this matched pair belonging to TAG, CE and PG classes were highlighted. *(b)* Lipid*ome* map of the matched ADC-tumor tissue of ID64. *(c)* Hierarchical clustering of 43 tissue lipid*omes* (subset training, see Tab. 4.3) based on LUX scores. Lipid*omes* are color coded according to the tissue classification below the dendrogram: gray: alveolar tissue; orange: ADC tissues; blue: SCC tissues; other: combines an LCC and a sarcomatoid LCC. The frequency of reoccurrence of each branch after 100 error modelling iterations is indicated with green numerals for a detection threshold (*dt*) of 0.005 and a standard deviation (SD) of 0.002. Red stands for a *dt* of 0.003 and a SD of 0.001. Modified from Eggers *et al.* [146].

ADC-tumor lipid*ome*, specifically 33 TAG species and 16 CE species were detected which were not present in the corresponding alveolar tissue. These remarkable lipid*ome* alterations were observed in a variety of paired tissue samples, suggesting that widespread lipid*ome* compositional changes occurred in the tumor lipid*ome*.

A hierarchical clustering was calculated on the basis of LUX scores between each pair of lipid*omes* to analyze their homology (Fig. 2.5, c). There was a stable tree morphology, which was shown by counting the frequency of reoccurrence of the branching structure by error modeling. The tree had two main branches derived as clusters. The first cluster (Fig. 2.5, c, left cluster) was exclusively composed alveolar lipid*omes*, including ID64. The second cluster con-

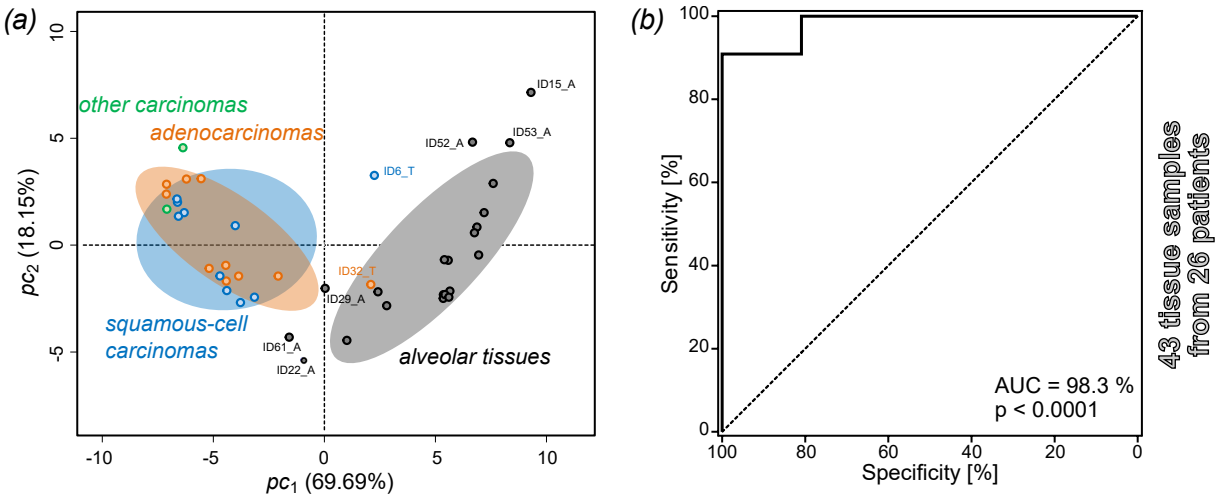


Figure 2.6: **Tissue discrimination based on LUX scores.** *(a) pca* based on pairwise LUX scores as input data (same input data as for Fig. 2.5). *pca* factor maps shows pc_1 and pc_2. *(b)* ROC curve based on pc_1 as discriminator for alveolar and tumor tissue lipid*omes*. Modified from Eggers *et al.* [146].

tained all tumor lipid*omes* and a sub-cluster of six alveolar lipid*omes* indicating that these alveolar lipid*omes* had a higher structural similarity to the tumors than to the other alveolar tissues. No pairs of tumor and alveolar tissues from the same patient clustered together, suggesting that a patient-independent tumor lipid*ome* has structural attributes present only in tumors *or* alveolar tissues.

In parallel, LUX scores were analyzed by *pca*. A clear separation of tumor and alveolar tissues along pc_1 was observed (Fig. 2.6, a). The axis pc_1 alone accounted for 69.69 % of the variability of the LUX score data was suitable as a discriminator of tumor and alveolar tissue lipid*omes*. A receiver operating characteristics (ROC) curve based on pc_1 had an area under the curve (AUC) of 98.3 % (Fig. 2.6, b) which proved excellent diagnostic properties of this methodology.

Taken together, the lipid*ome* analysis based on LUX scores demonstrated that alveolar and tumor tissues had different lipid*ome* compositions. This indicated general lipid metabolic changes between these types of tissue. The limitation of this methodology was shown by a more detailed inspection of the results for the tissue phenotype, *e. g.* the tumor entity. With this LUX score based analysis it was not possible to distinguish between ADC and SCC tissues. The reason for this observation could be that the structural information is only one part of the lipid*ome*. No doubt, the structural information is important and often ignored. However, the integration of the quantitative data is not negligible. The quantitative part of the lipid*ome* has not yet been

used in the calculation of LUX scores. In order to complete the image, quantitative information could be integrated into this methodology. Potentially, this would open great opportunities for in-depth lipid*ome* analyses.

2.2.2 Systematic alterations between NSCLC and alveolar lipid*omes*

LUX scores revealed that alveolar and tumor lipid*omes* had specific molecular compositions. Further analyses of the quantitative lipid*ome* characteristics in lung tissues were performed and an unsupervised hierarchical clustering was calculated. Thus, 141 lipids, quantified in at least 90 % of the samples were used for this analysis. The hierarchical clustering dendrogram showed a clear separation of alveolar tissues on the left and tumor tissues on the right (Fig. 2.7) which nicely confirmed the results obtained by LUX score analysis with respect of individual lipid*ome* features in alveolar lung tissues and tumors. All alveolar tissues, except ID29_alveolar tissue (A) who suffered from Crohn's disease, formed one cluster with a correlation coefficient of 0.779 (Fig. 2.7, b). The tumor lipid*omes* showed a higher heterogeneity compared to the alveolar tissues as indicated by the branching structure. Tumors formed three sub-clusters with low correlation coefficients to each other (Fig. 2.7, b). ID24_tumor tissue (T) separated from other tumors, possibly due to the fact that this tissue was completely necrotic.

The neutral lipids DAG, CE and TAG were present in higher amounts in the tumor tissues compared to the alveolar tissues (Fig. 2.7, b). Furthermore, all TAG and CE species showed similar behavior across all samples, indicated by a correlation of those species with a coefficient of 0.781 on the tree of the lipids (Fig. 2.7, a). In contrast, major components of the pulmonary surfactant were present in high amounts in alveolar tissues whereas these lipids were barely recognized in tumors. This was most strongly observed for PG species but also for the saturated GPLs PC 32:0, PC 30:0 and PC 28:0. This finding might indicate that a general loss of surfactant occurred during carcinogenesis. All PG lipids had similar behavior across all samples which was demonstrated since they formed a separate branch in the dendrogram for the lipids (Fig. 2.7, a). Another observation was that long-chain SM species as well as PS' had higher concentrations in the alveolar tissues. Unfortunately, no systematic differences were identified between lipid*omes* of ADC and SCC tissues.

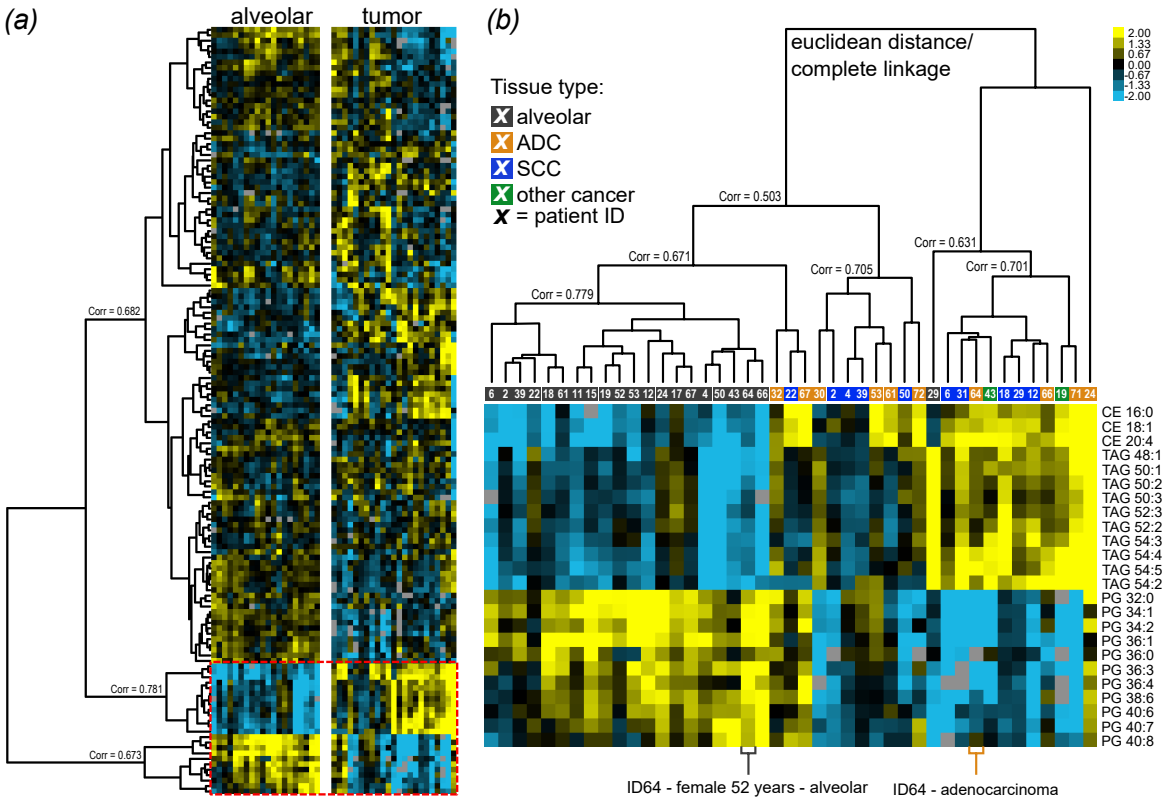


Figure 2.7: **Lipid*omes* of alveolar tissues and tumors had distinct lipid profiles.** *(a)* Heat map and hierarchical clustering summarizing 141 lipid species quantified in 43 tissue samples. *(b)* Dendrogram for the tissue lipid*omes*. Lipid*omes* are color coded according to the tissue classification below the dendrogram: gray: alveolar tissue; orange: ADC tissues; blue: SCC tissues; other: combines an LCC and a sarcomatoid LCC. Modified from Eggers *et al.* [146].

2.2.3 The histopathological phenotype is reflected in lipid*omes* of NSCLC tissues

The heterogeneity in the tumor lipid*omes*, as observed by the hierarchical cluster analysis, indicated a complex pattern of unknown parameters that are influencing the tissue lipid*omes*. These could include the tissue morphology as well as patient related parameters such as obesity, metabolic diseases or the lifestyle like smoking or physical activity. Based on this observation, PLS regression analysis was performed to identify parameters that affect the tumor tissue lipid*ome* composition. The PLS regression model combined the quantities of 113 lipids that were present in all 22 tumor tissue samples with clinical and histopathological data.

A basic assumption was that the histopathological phenotype caused significant lipid*ome* alterations. To verify this hypothesis, lipid*omes* of the tumors were compared to a description of

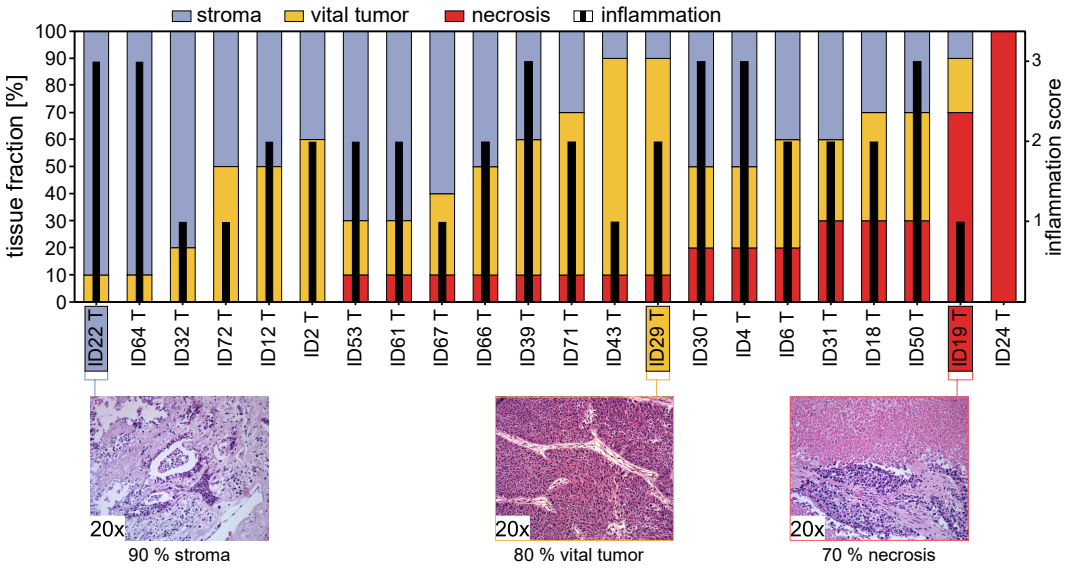


Figure 2.8: **Histopathologic scoring of tumor tissues showed a great heterogeneity.** Results of histological scoring of the pilot study. Sections were histologically characterized according to their tissue composition, using percentages of necrotic areas, vital tumor cells and stroma content. The infiltration by immune cells (inflammation) was scored in categories (0, 1, 2, 3). Modified from Eggers *et al.* [146].

the histopathologic phenotype by scores which reflected the tissue morphology. Characteristics of the tissue composition were expressed as tissue fraction which indicated the percentages of necrotic areas, stroma and metabolically active tumor cells. In addition, the inflammation status was categorized in stages from 0 (no inflammation) to 3 (highest inflammatory status) (Fig. 2.8). The overview of the determined histology scores of tumors (Fig. 2.8), sorting the tissues according to their necrosis content, showed a large heterogeneity. Especially the examples with 90 % stroma (ID22_T), 80 % vital tumor (ID29_T) and 70 % necrosis (ID19_T) showed the large morphological alterations expected to be reflected in the lipid*omes*.

Evaluation of the PLS regression factor map (Fig. 2.9, a) demonstrated that it was indeed the tissue fraction that had the most potent effects on the tumor lipid*omes*. On the PLS factor map, which reflects the similarity of the lipid*omes* within the parameters of the models, the tumor lipid*omes* sorted according to the tissue fraction (Fig. 2.9, a). Lipid*omes* from samples with high stroma content were localized in the top-left corner of the factor map and were further separated from samples having a high proportion of vital tumor along the PLS component t_2. As consequence, samples with high vital tumor content localized at the bottom of the factor map. There were two highly necrotic samples located at high t_1 values (right side of the factor map). The three parameters, making the tissue fraction, had markedly different characteristics within

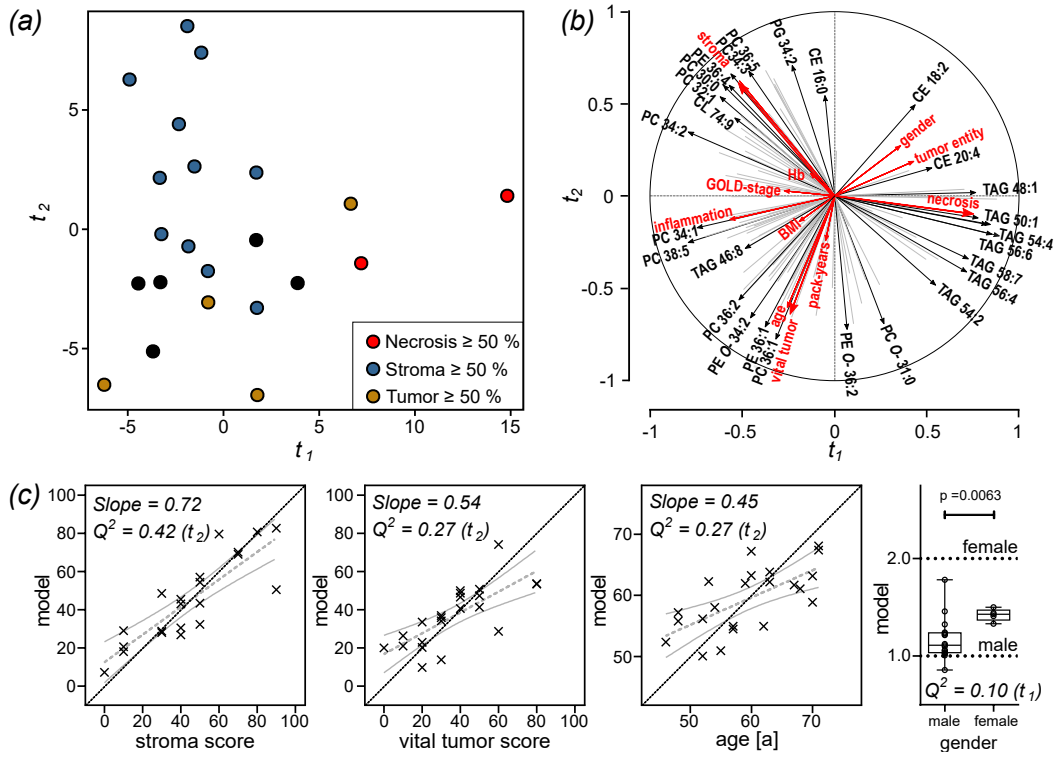


Figure 2.9: **The histopathological phenotype is reflected in the lipid*omes* of tumor tissues.** *(a)* Factor map of individual tumor lipid*omes*, color coded according to the dominant tissue fraction (stroma >50 %, blue; vital tumor cells >50 %, yellow; necrotic areas >50 %, red; none of the tissue fractions >50 %, black). *(b)* Correlation of lipid quantities, histology scores, and clinical data to the *t*-components. The vectors indicate how strong the variables correlate to the *t*-components and show the correlations between the lipid*ome* data (black arrows) and the histology scores and clinical data (red arrows). *(c)* Correlation of original histology scores to scores computed from the PLS regression model for stroma and vital tumor as well as correlation for the clinical parameters age and gender. Linear regressions with 95 % confidence bands are shown in grey. The specific slope is noted on the plot. The dotted black line indicates the location of the ideal correlation. The boxplot shows computed scores versus the original categorical values. Q^2 values are the results of the cv of the model. A model is considered significant for a response if Q^2 is > 0.0975. The associated *t*-component is noted in brackets after the Q^2 value. The *p*-value was calculated using the U test. Modified from Eggers *et al.* [146].

their lipid*omes*. This was indicated by the correlations of the variables and the *t*-components (Fig. 2.9, b) in which vectors of tumor, necrosis and stroma are in opposite directions. This revealed that each of these parameters had unique features in its lipid*ome* which were independent to the others. The stroma content correlated most strongly with lipids such as PC 36:5 and PC 34:3 as well as with pulmonary surfactant related lipids PC 32:0 and PC 30:0. The vital tumor content showed correlations with the GPL species PC 36:1 and PE 36:1, and necrosis was associated with TAG molecules (Fig. 2.9, b).

To test the results for their stability, to avoid overfitting the models and to validate predictive power, a cv process was performed. Significant correlations were identified for the stroma and

vital tumor scores with cross-validated regression coefficient (Q^2) values of 0.42 and 0.27, respectively (Fig. 2.9, c). Necrosis did not reveal a significant correlation, possibly since only two samples had more than 50 % necrosis. Aging appeared to affect tumor lipid*omes*, which was characterized by a significant cv ($Q^2 = 0.27$) (Fig. 2.9, c). From the available data, a possible dependence between aging and vital tumor content cannot be excluded, since a correlation of both parameters was observed (Fig. 2.9, b). For the categorical parameters, such as gender, the predicted score of the regression model was compared with the original categorization (Fig. 2.9, c). A U test was then calculated to test whether the score distinguished between the categorizations. For gender, a significant regression model (Q^2 = 0.1) was calculated which had enough predictive power to distinguish between tissues from males and females ($p = 0.0063$) (Fig. 2.9, c). Other parameters that have been tested in this regression model, *e. g.* the tumor entity, the inflammatory stage, the hemoglobin (Hb) content as a measure for residual blood in the tissues, pack years (PY), BMI and stage of COPD according to the global initiative for chronic obstructive lung disease (GOLD) [147] did not show any significant models.

2.2.4 Emphysema and aging reflected specific lipid*ome* alterations of alveolar lung tissues

The regression analysis calculated from the lipid*omes* of the tumor tissues demonstrated that the histopathological phenotype was mirrored in the lipid*omes*. The same approach was used to evaluate factors potentially influencing the alveolar lipid*omes*. The histopathological phenotype was more homogenous for alveolar tissues when compared to tumors (Fig. 2.7, 53). The cell density was observed to be highly variable and was for instance expressed by the pulmonary emphysema score. In which manner the phenotype changes with respect to enlarged alveolar structures and lower cellular density was shown by the example tissues from the stages 2 (ID39_A) and 5 (ID2_A) (Fig. 2.10). Due to this importance of emphysema, the overview of scoring results was sorted by the extent of emphysema (Fig. 2.10). Furthermore, the inflammation status, density of AM and pulmonary fibrosis were evaluated and categorized in histopathologic scores from 1 to 3 (Fig. 2.10).

Based on the histology scores and clinical parameters, a PLS regression model was calculated including 139 lipid species that were quantified in all alveolar tissue lipid*omes*. The PLS

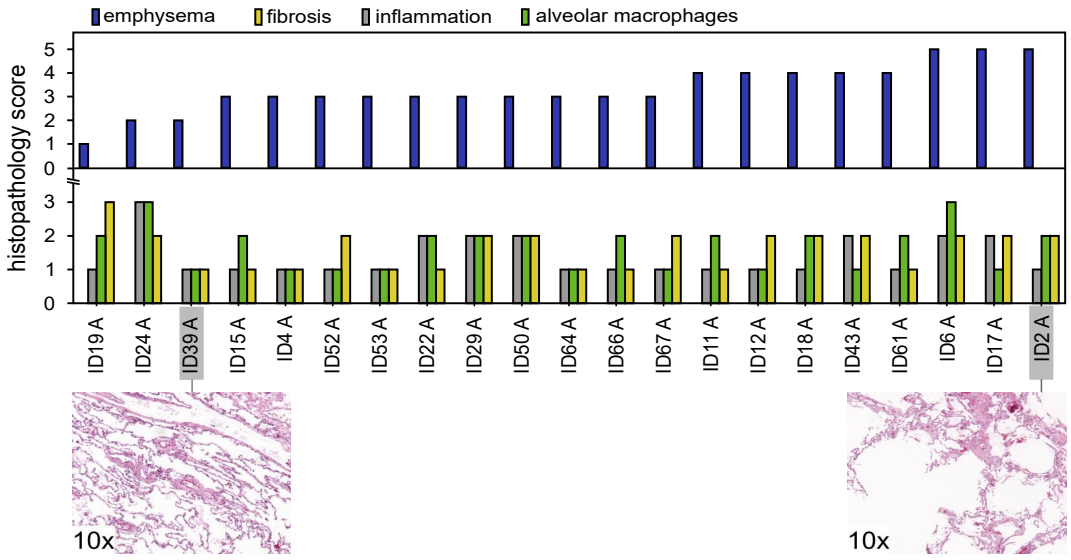


Figure 2.10: **Histopathological characterizations of alveolar tissues reflect large heterogeneity.** Alveolar tissues were scored by inflammation (stages 0, 1, 2, 3), AM (stages 0, 1, 2, 3), emphysema according to Thurlbeck [148] and fibrosis (stages 0, 1, 2, 3). The bar graphs show the scoring results for every individual tissue, sorted according to increasing emphysema score form left to right. Modified from Eggers *et al.* [146].

regression factor map showed decreasing emphysema scores with increasing t_1 values since the lipid*omes* sorted according to the emphysema parameter from right to left (Fig. 2.11, a). The correlation circle further indicated that emphysema and age correlated with each other (Fig. 2.11, b). This pilot study showed that emphysema and aging positively correlated with the lipid species PS 38:4, PI 38:4, PI 38:5 and PS 40:4. The cv procedure resulted that emphysema and aging were significant on the component t_1 with Q^2 values of 0.16 and 0.14, respectively. These results indicated that aging and emphysema could not completely be separated but they might induce systematic lipid*ome* alterations. This fits to earlier observations that aging and progression of COPD and emphysema are potentially be linked [149].

The PLS regression model indicated that the classification of the carcinoma (tumor entity) might be reflected in the lipid*omes* of tumor-free lung tissues (Fig. 2.11, c). At this stage potential limits in utilization of residual material from lung cancer patients for investigation of respiratory diseases were reached when lung cancer induces specific lipid metabolic alterations even in the tumor-free compartments of the lung. Especially, patients suffering from large-cell carcinomas (LCCs) and carcinoid tumors (grouped as others) were distinguished from patients with ADC and SCC diagnosis with p-values of 0.0167 and 0.022 (Fig. 2.11, c). Due to the small number of cases (21 alveolar tissue) potential overlapping effects with other parameters

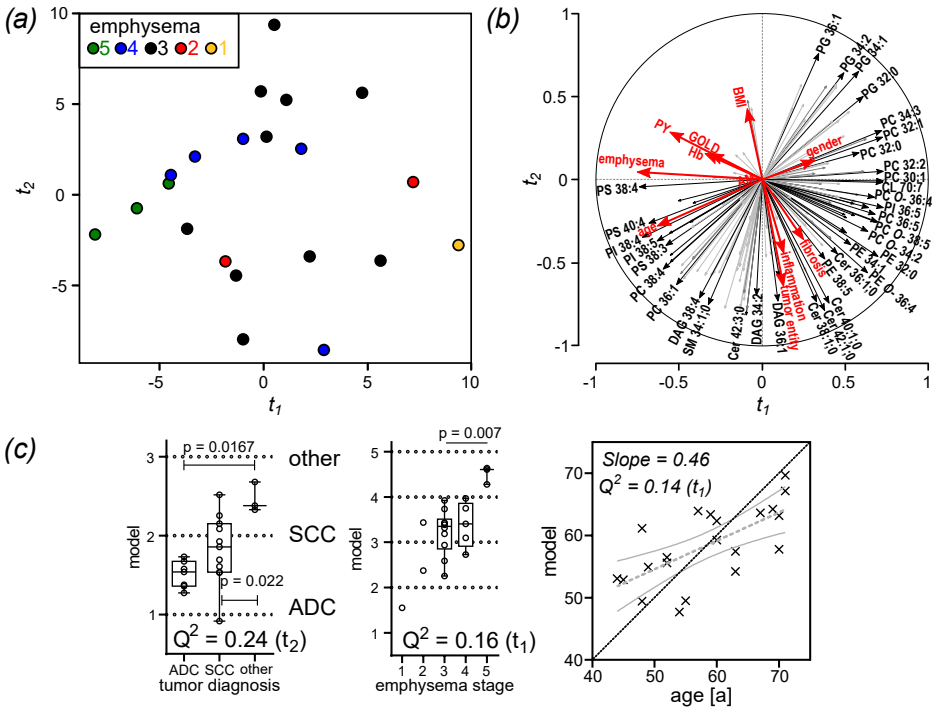


Figure 2.11: **Parameters affecting lipid*omes* of alveolar lung tissues analyzed by partial least-squares regression.** *(a)* PLS regression factor map for alveolar tissue lipid*omes*. Individuals were color coded according to their emphysema score. *(b)* Correlation of lipid quantities, histology scores and clinical data to the *t*-components. Predictor variables are indicated by black vectors and responses by red vectors. *(c)* Correlations of the PLS regression model between original values and computed values for tumor entity, emphysema stage and age. Grey dotted lines represent the linear fit, and solid grey lines indicate 95 % confidence bands. The dotted black line indicates the location of the ideal correlation. Q^2 values are the results of the cv of the model. A model is considered significant for a response if Q^2 greater than 0.0975. The associated *t*-component is noted in brackets after the Q^2 value. All *p*-values were calculated using the U test. Q^2 values and *p*-values are only noted when significance was reached. Modified from Eggers *et al.* [146].

with regard of inflammation status and fibrosis were not excluded. Smoking behavior could be connected to the histological tumor entity. It is known that SCC and ADC, both, occur predominantly in smokers. However, the ADC alone has high prevalence in never-smokers [80, 81]. For the remaining parameters inflammation stage, fibrosis, BMI, PY, gender, Hb content and GOLD stage of COPD no significant correlations were observed from the regression model.

Taken together, the PLS regression model nicely reflected that emphysema and aging are parameters shaping lipid*omes* of alveolar lung tissues. However, the number of patients is rather limited for further conclusions and closer functional associations with regard to metabolic pathways. Therefore, it was necessary to increase the numbers of cases for deeper investigations.

2.3 The lipid*ome* mirrors a complex pattern of clinical parameters and pathological events

Results from the pilot study were compared to a follow-up study which was performed with an increased number of cases. The tissues collected in *sub-cohort 2* were combined with *sub-cohort 1* that tissues from 92 patients were available for this follow-up study. Lipid profiles of alveolar lung tissues and tumors were investigated for specific patterns appearing during carcinogenesis. It was asked whether the lung cancer entity was reflected in the lipid profiles. An advanced PLS regression modeling approach was utilized which combined lipid*omics* data with histopathological and clinical parameters and provided an overview which of these parameters had the highest influences on the lipid*omes* of lung tissues.

2.3.1 Heterogeneous lipid profiles in alveolar and tumor tissues

A hierarchical cluster analysis was performed with the lipid*ome* data from this enlarged cohort. For correlation analyses, quantities of 145 lipid species, present in at least 80 % of the 174 tissue lipid*omes*, were used as input. Hierarchical clustering of tissue samples indicated a complex branching structure comprising 6 clusters (Fig. 2.12). The most frequently populated clusters were the clusters 4 (94 tissue lipid*omes*) and 5 (69 tissue lipid*omes*) which revealed clear separation of alveolar and tumor tissue lipid*omes*. Cluster 4 contained 90 % alveolar tissues but also a small subset of 10 % tumor tissues, mostly ADCs. This indicated that there was a low number of tumor tissues that was closer related to the alveolar tissues than to the tumors. Cluster 5 comprised 97 % tumor tissues and 3 % alveolar tissues (ID126_A and ID104_A). Cluster 1 contained only the necrotic ADC tissue ID24_T and separated from all other tissues as it was observed before in the pilot study (Fig. 2.7, page 53). The clusters 2 and 3 contained one or two lipid*omes* each, mainly tumors who were low correlated to the other clusters. Cluster 6 separated from the other lipid*omes* and comprised lipid*omes* from carcinoid tumors and one ADC tissue.

All clusters were characterized by a specific pattern within their lipid profiles. Cluster 4 (alveolar tissues) showed an overall decrease of the neutral lipids TAG and CE. In cluster 5 (tumors), abundances of CE species such as CE 18:1 and CE 20:4 were overall increased. Moreover,

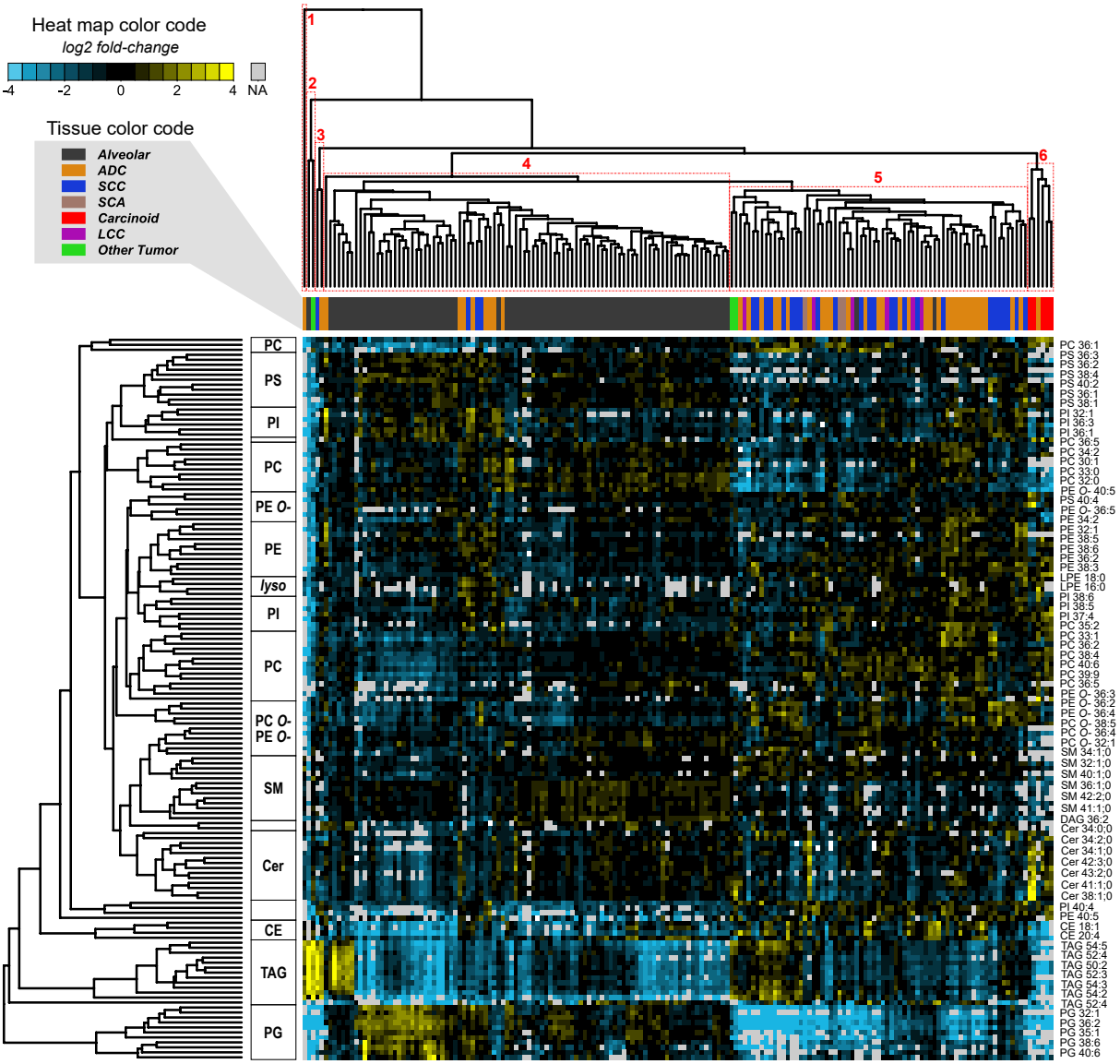


Figure 2.12: **Tumor and alveolar tissues were distinguished by their lipid profiles.** Hierarchical clustering was calculated with 145 lipids that were present in at least 80 % of the used 174 tissue samples. The heat map shows log2 fold-changes compared to the overall mean value across all tissues for the respective lipid species. Hierarchical clustering was calculated using Euclidean distance as similarity metric and average linkage as clustering method. Clustering was performed for lipid species as well as for the tissue samples. On the dendrogram for tissue lipid*omes*, 6 clusters were observed as indicated by red numbers. Classification according to the tissue type is annotated below the tree in a color code. Lipid classes are annotated left to the heat map indicating the main classes and each second lipid species is annotated right to the heat map.

cluster 5 (tumors) showed both, a sub-cluster with decreased abundances of TAGs, and a small cluster that had slightly increased levels of them. The clusters 2 and 3 (mainly tumors) revealed strongly increased TAG levels. It was already recognized in the pilot study that lipid*omes* of alveolar tissues had high levels of PG lipids whereas these species were low concentrated in the tumor lipid*omes* (section 2.2.2, page 52). This observation could be specified by closer inspection of the heat map (Fig. 2.12). A large number of PG species which also included the surfactant main components PG 34:1 and PG 36:2 was low concentrated in the tumors (clusters 2, 3, 5 and 6) compared to the alveolar tissues (cluster 4) which was shown by log2 fold-changes of -4. However, a high variability within the measured abundances of PGs with the subset of alveolar tissues was recognized which potentially indicated development of lung pathologies there.

All these findings were confirmed by *pca*. A clear separation of tumor and alveolar tissues was accomplished on the factor maps along pc_1 where alveolar tissues had negative values and and tumors positive (Fig. 2.13, a and c). Carcinoid tumors clustered outside of the NSCLC tissues and had high values of pc_1 (Fig. 2.13, a and c). The axes pc_2 and pc_3 did not separate ADC and SCC tissues. The surfactant associated lipids PG 36:1, PG 34:2 and PG 30:0 negatively correlated to pc_1 indicating increased amounts of those lipid species in alveolar tissues (Fig. 2.13, a and b). Long-chain and unsaturated PC and PE species such as PC 40:6, PC 36:2, PC 38:5 and PE 40:5 correlated positively with pc_1 which indicated an overall increase of those species in tumors. The distribution of TAGs, SMs and ether lipids did not correlate with the tissue type. Abundances of TAG species such as TAG 56:6 and TAG 54:2 negatively correlated with pc_2 (Fig 2.13, b). This observation was in-line with the heat map (Fig. 2.12) where the TAG levels were not associated to the tissue classification which stands in contrast to the results from the pilot study. Potentially, there were other parameters that associated with TAGs than the tissue classification. Abundances of ether lipids such as PC *O*- 36:5 and PC *O*- 38:5 as well as long chain SM species such as SM 42:2;0 and SM 42:3;0 positively correlated with pc_3 (Fig. 2.13, d).

A larger number of cases further increased the complexity of the branching structure of the hierarchical clustering. This was shown by the fact that the clear separation of alveolar and tumor tissues, as observed in the pilot study, was not accomplished. Therefore, another approach was used in the next step to extract specific features which altered with the tissue phenotype.

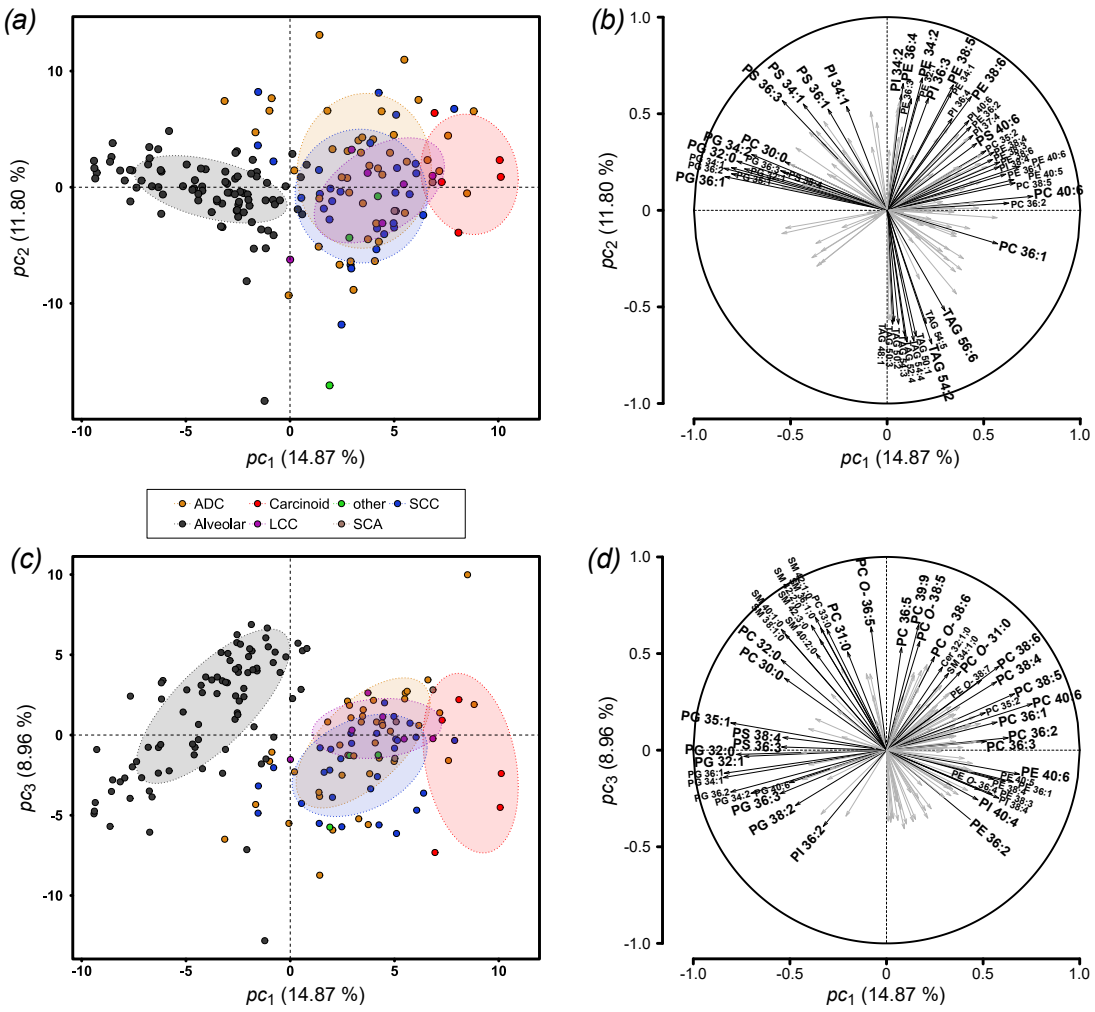


Figure 2.13: **Principal component analysis confirmed observations made by hierarchical clustering.** *(a) pca* factor map. The same input dataset as used for figure 2.12 comprising 145 lipids from 174 tissue samples was used. Factor map shows pc_1 and pc_2. Percentages of included variance are noted in axis-labels. Each dot on the map represents one tissue lipid*ome*. The dots are color coded according to their tissues type as noted in the legend. Ellipses show confidence areas including 68 % of all points per group. *(b)* Circle of correlations from the same analysis of pc_1 and pc_2 and the variables (lipids). The longer a vector, the better the variable is represented in the analysis. Directions give the component the variable correlates. The 50 most contributing variables are annotated. *(c)* Factor map from the same analysis showing pc_1 and pc_3. *(d)* Circle of correlations for pc_1 and pc_3.

2.3.2 Specification of a lipid panel to differentiate between alveolar and tumor tissues

Lipid*omes* of alveolar and tumor tissues were compared to identify a specific lipid panel which was able to differentiate between them. Utilizing unpaired U tests, a panel of 39 lipid species was identified which were significantly changed between both groups and had a log2 fold-change greater than 1 or lower than -1 (Fig. 2.14, a). The observation of increased PG levels in the alveolar tissue lipid*omes* as made by the unsupervised correlation analyses (Figs. 2.12 and 2.13, pages 60 and 62) was confirmed. Specifically, surfactant lipids such as PG 36:2, PG 32:0, and PG 34:1 as well as lower abundant PG species including PG 40:7 and PG 35:1 were elevated in the alveolar tissues. Theses lipids clustered outside from the other selected lipids as shown by the hierarchical clustering (Fig. 2.14, d, area marked with II). The lipid panel included three SM species (SM 40:1;0, SM 38:1;0 and SM 41:1;0) that were increased in the alveolar tissues (Fig. 2.14, a). Distribution patterns of these species correlated with the surfactant main components PC 32:0 and PC 30:0 (Fig. 2.14, d, area marked with I). It was interesting that the PCs did not correlate with the PGs although both were known to originate in the pulmonary surfactant. This observation indicated that the surfactant composition was not constant through all the samples. Potentially the PC-to-PG ratio in the tissue lipid*omes* was an indicator for lung pathologies. The second interesting finding was that the long-chain SM species correlated with surfactant PC. SM species are known to be present in lipid rafts of apical membranes [46]. Potentially, SMs are markers for active alveolar surface together with the pulmonary surfactant.

Tumor tissues had elevated levels of TAGs and CEs as it was already noticed in the pilot study (Fig. 2.7, page 53). Long-chain and highly unsaturated as well as ether lipids, such as PE 40:5, PE 40:6, PI 40:4 and PE *O*- 36:3 were overall increased in tumors. The ether lipids have also been observed with increased abundances in tumors in the cluster analysis (Figs. 2.12 and 2.13).

The next question was whether the selected lipid panel was suitable as a diagnostic tool. Therefore, the lipid abundances were used to calculate discriminative scores for each tissue sample (Fig. 2.14, b). The result was that alveolar tissues had scores of about -50 arbitrary units (*au*) whereas tumors had scores of about +50 *au* (Fig. 2.14, b). Afterwards, the predictive power of the score was tested in a ROC curve which plots the true-positive-rate (sensitivity)

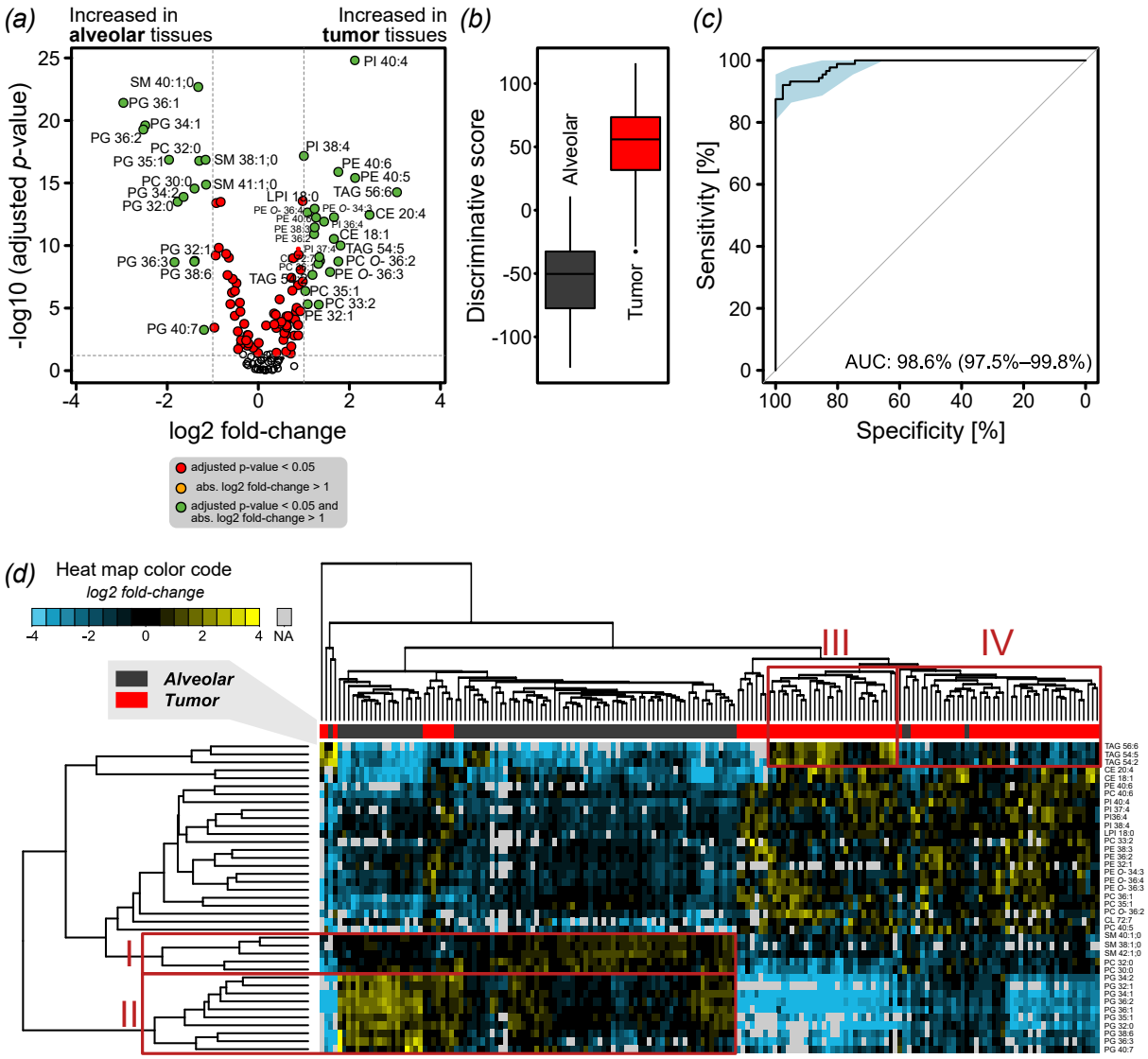


Figure 2.14: **Alveolar and tumor tissues can be differentiated by a lipid panel.** *(a)* Volcano plot showing significantly changed lipid quantities between tumor tissues and alveolar tissues. The *p*-values were calculated by unpaired U tests with Benjamini-Hochberg FDR. Lipid species with adjusted *p*-values lower than 0.05 are shown in red. Yellow indicates lipids with log2 fold-changes of group-wise mean values less than -1 or larger than 1. Green dots represent lipids with *p*-values lower 0.05 *and* log2 fold-changes less -1 or larger than 1. Lipids drawn on the left were significant for alveolar tissues and lipids drawn on the right side for tumor tissues. *(b)* Box plot of the discriminative score. The score was based on lipid abundances of the green marked species from panel a. The discriminative score was calculated by multiplication of each lipid abundance value with the respective log2 fold-change and calculation of the sum across each sample. *(c)* The ROC curve for the differentiation between tumor tissues and alveolar tissues based on the discriminative score. The blue area represents the 90 % confidence area. *(d)* Hierarchical clustering and heat map representation of the selected lipid panel. On the heat map, lipid abundances were expressed as log2 fold-changes compared to the overall mean per lipid species. Hierarchical clustering was calculated using Euclidean distance as metric and average linkage as clustering method. The color bar below the samples tree indicated the tissue type. Gray marks alveolar tissues, red tumors.

over the false-positive-rate (specificity) of a predictive score. The ROC curve showed high specificity with high sensitivity which confirmed a great diagnostic potential (Fig. 2.14, c) (AUC = 98.6 %).

A hierarchical clustering based on the selected lipid panel showed a clear-cut separation of tumors and alveolar tissues (Fig 2.14, d). However, a low number of tumors clustered between alveolar tissues and *vice versa*. The main cluster of tumor tissues was subdivided into two sub-clusters (Fig. 2.14, d, areas marked with III and IV). Sub-cluster III was characterized by increased levels of TAG 56:6, TAG 54:5, TAG 54:2, CE 18:1 and CE 20:4. In contrast, the sub-cluster IV had lower levels of the TAGs. A set of four tissue lipid*omes* (ID24_T, ID125_T, ID100_A and ID65_T) clustered outside of all other tissues. These lipid*omes* were characterized by high abundances of the TAG species. ID24_T, ID125_T, and ID65_T were striking in the correlation analyses before (Fig. 2.12). ID24_T was necrotic and histopathology showed that ID125_T was a sample from a lymph-node accidently sampled as tumor. For ID65_T and ID101_A no histopathological characterizations were available.

Overall, the presented findings demonstrated that lipid profiles were sufficient to discriminate lung cancer from normal lung tissues. To learn more about lipid metabolic alterations in specific tumor entities, *e. g.* differentiating ADC from SCC tissues, subsets from the initial dataset were analyzed which reduced the number of lipid*ome* influencing parameters with setting one parameter constant. This allowed exploring specific signatures for different tumor entities.

2.3.3 Identification of lipid*ome* alterations between tumor entities

The next question was whether the analysis approach of section 2.3.2 was able to identify lipid panels that were suitable for discrimination of specific tumor entities. Therefore, ADC tissue lipid*omes* were compared with SCCs and NSCLC lipid*omes* were compared with LCCs. Within each pairwise comparison of these subsets, lipids were included which were present in at least 97 % of all the compared samples. It was necessary to allow not more than 3 % missing values per lipid species since groups compared were not equally sized such as ADC (N = 41) and LCC (N = 6). There, an occupation cut-off of 80 to 90 % resulted in datasets where the smaller group had too less or no observations. ADC tissues were the most frequent fraction within the tumor cohort (47 %) followed by SCC (32 %), LCC (7 %) and carcinoid tumors (6 %). The

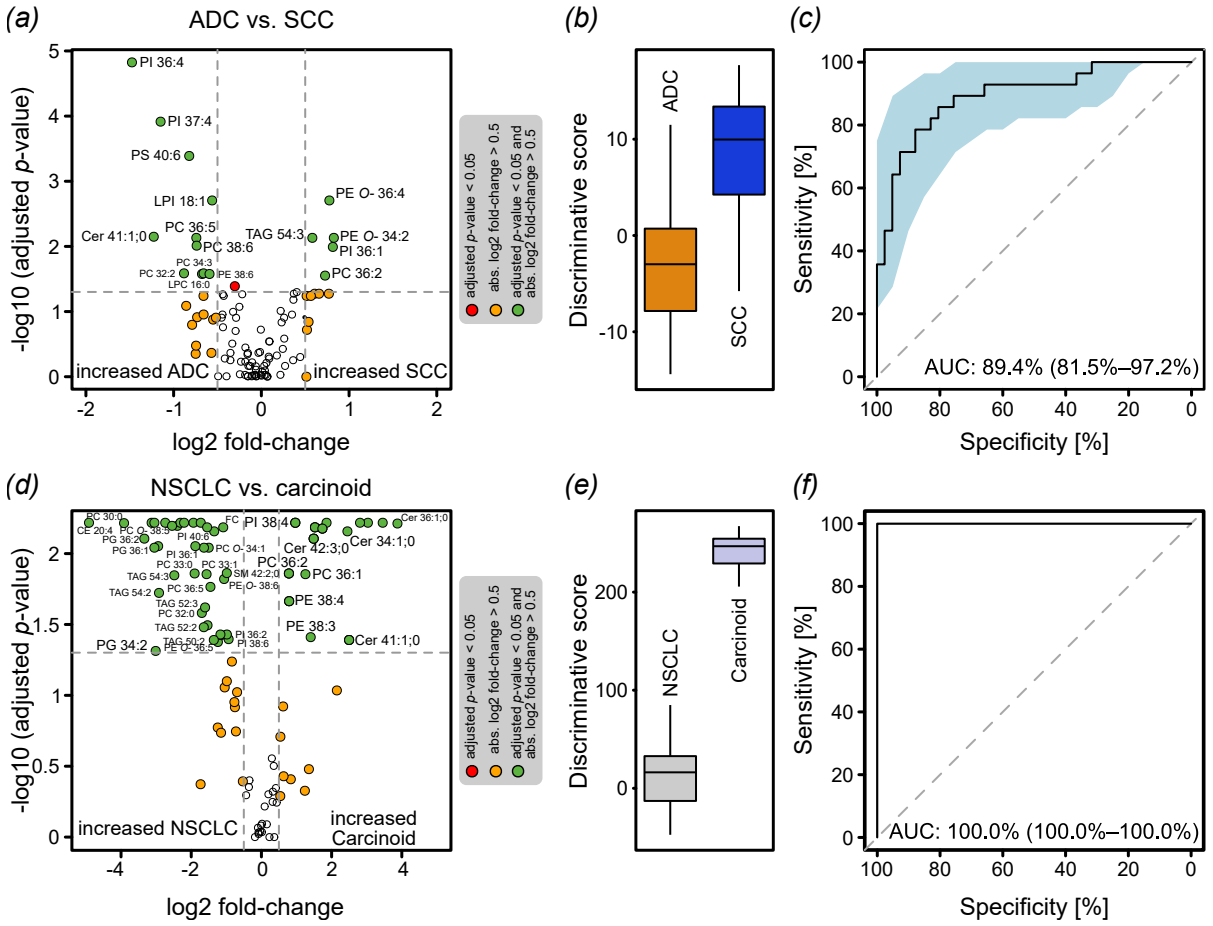


Figure 2.15: **Identification of lipid panels for discrimination of different tumor entities.** *(a)* Volcano plot comparing ADC (N = 41) and SCC (N = 28) lipid*omes*. *(b)* Boxplot of the score discriminating ADC and SCC lipid*omes*. The score is based on the green marked lipids in panel a. The score was calculated in the same manner as noted for figure 2.14. *(c)* ROC curve based on the discriminative score. The blue area represents the 90 % confidence area. *(d)* Volcano plot comparing NSCLC (N = 75) lipid*omes* (ADC, SCC, and LCC) with carcinoid tumor lipidomes (N = 5). *(e)* Discriminative score for NSCLC vs. carcinoid. *(f)* ROC curve based on the discriminative score for NSCLC vs. carcinoid.

absolute log2 fold-change cut-off for lipid selection was set to 0.5 to include a larger number of lipid species in the discriminative score.

The U tests revealed no significantly altered lipid species between ADC and LCC tissue lipid*omes*. ADC and SCC tissue lipid*omes* distinguished from each other by a panel of 14 lipids (Fig. 2.15, a). Amounts of long-chain and highly unsaturated lipids such as PI 36:4, PS 40:6, PC 36:5 and PC 38:6 increased in the ADC tissues. Ether lipids such as PE *O*- 36:4 and PE *O*- 34:2 together with *e. g.* PI 36:1 and PC 36:2 increased in the SCC tissues. The magnitude of changes was comparable low, yielding in absolute log2 fold-changes below values of 1. High levels of significance were reached with p-values down to 10^{-5} for PI 36:4. These p-

values indicated highly specific alterations between both lung tumor entities. The discriminative score based on the selected lipid panel had values about -3 *au* for the ADC lipid*omes* and values about 10 *au* for the SCC lipid*omes* (Fig. 2.15, b). The corresponding ROC curve revealed that the score was suitable as diagnostic tool (AUC = 89.4 %, Fig. 2.15, c).

The hierarchical clustering and *pca* had shown that carcinoid lipid*omes* separated from the other tumors (Figs. 2.12 and 2.13). To investigate the specific lipid*omes* of carcinoid tumors, these were compared with NSCLC tissues (combination of ADC, SCC and LCC) (Fig. 2.15, d - f). A panel of 54 lipids was identified which had altered abundances between carcinoid and NSCLC lipid*omes*. The *p*-values were larger than 10^{-3} and log2 fold-changes had values between -4 and 4 (Fig. 2.15, f). The carcinoid tumors showed increased Cer levels including the species Cer 34:1;0, Cer 36:1;0, and Cer 42:2;0 whereas the corresponding SM species SM 34:1;0 and SM 42:2;0 were significantly decreased which could indicate alterations in the SL metabolism. A number of GPLs was decreased in the carcinoid lipid*omes* which included PI 38:4, PI 38:5, PC 36:3, PC 36:2, PC 36:1, PC 38:4, and PE 38:3. Also, the neutral lipid classes CE and TAG with species such as CE 18:1, CE 20:4, TAG 54:2 and TAG 52:2 were significantly decreased in the carcinoid tumors. Moreover, the carcinoid tumors had, compared to the NSCLC tissues, decreased abundances of surfactant associated lipids including the prominent species PG 36:2, PG 36:1 and PC 32:0. This indicated a general loss of surfactant synthesis in carcinoid tumors. The discriminative score for carcinoid and NSCLC lipid*omes* had values around 0 *au* for NSCLC tissues and values larger than 200 *au* for carcinoid tumors (Fig. 2.15, e). The range of variation of the scores for NSCLC and carcinoid lipid*omes* did not overlap. Consequently, the score unambiguously discriminated carcinoid tumor lipid*omes* from NSCLC lipid*omes* (AUC = 100 %, Fig. 2.15, f). This result demonstrated, that carcinoid tumors have unique properties within their lipid metabolic processes which could be of high interest for further research.

These results showed in which manner tumor lipid*omes* reflected the histopathological characterization of the tissues. It was also indicated that a specific heterogeneity in the tumor samples regarding necrosis, vital tumor content and stroma influenced the tumor lipid*ome* composition. However, there were still manifold open questions how these models could be used for prediction and what are influencing parameters which could be for example located in the lifestyle of the analyzed patients.

2.3.4 An advanced PLS regression strategy identified tumor lipid*ome* influencing parameters

The pilot study revealed PLS regression as a powerful tool to evaluate parameters that affect the lipid*ome* composition of lung tissues (section 2.2.3, page 53). The cv procedure provided information about the significance of the resulted models. However, the number of tissues included in the analysis was relatively low and there was no incidence how robust the obtained models were.

Therefore, an advanced modeling strategy was applied since the larger number of cases made it possible to split the cohort into training and validation subsets. The question was with respect to the robustness of the computed models. Therefore, the original cohort was randomly split 100 times into training and validation subsets. This resulted in 100 randomly compiled datasets and for each of these models a cv procedure was applied. With this approach it was possible to obtain the robustness of the calculated models utilizing randomly distributed training subsets.

Consecutively, this advanced modeling strategy was applied to further investigate which parameters were influencing the tumor lipid*omes*. The models combined histopathological characterizations regarding the content of necrosis, stroma and metabolic active tumor cells as well as the stage of inflammation. As tissue related parameter, the Hb content in the tissues was determined and considered as measure for residual blood in the samples. The clinical records containing age, gender, and BMI were assessed. The goal was to achieve a weighting which of these parameters were major or minor influencing the lipid*ome* compositions and whether these factors were predictable from the lipid profiles.

Tumor lipid*omes* were influenced by histopathological characterizations

Sixty two tumor lipid*omes* with full histopathological characterizations and almost complete clinical records were analyzed using PLS regression. As already observed in the pilot study (section 2.2.3, page 54), the tumor samples had a large heterogeneity within their histopathology (Fig. 2.16, b, blue). Vital tumor content and stroma were distributed widely from 0 to 100 % (Fig. 2.16). Most tissues had a necrosis content below 20 %. Higher necrosis proportions were underrepresented which had to be considered while interpretation. The inflammation status was approximately equally distributed in the stages 1, 2 and 3 (Fig. 2.16). The most frequent tumor

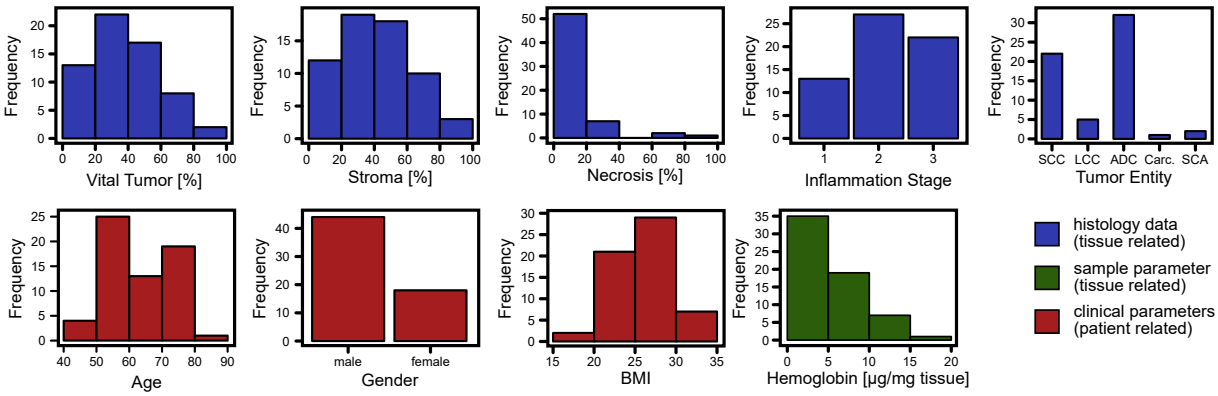


Figure 2.16: **Histopathological and clinical parameters that were associated to the tumor lipid*omes*.** Histograms show the distribution across all analyzed tumor samples. In blue, histopathological parameters are shown. In green noted is the sample parameter Hb content as measure for residual blood within the tissues. Red histograms show included clinical parameters related to the patient. Carc.: carcinoid.

entities were ADCs and SCCs. LCCs, carcinoids and sarcomatoid tumors (SCAs) were included as a minor subsets. Most of the included tissues contained less than 10 µg/mg Hb whereas a minor fraction had Hb levels up to 20 µg/mg. Patients ages ranged from approximately 40 to 90 years where the major portion was between 50 and 80 years old. Two third of the tissues originated from male patients and their BMIs ranged from approximately 15 to 35.

All these parameters were used as response variables in the PLS regression models. The lipid*omes* used as input data contained 127 lipid species that were present in at least 90 % of the 62 tumor tissues. Subsets were randomly created to contain 31 lipid*omes* each. Based on these 100 times 31 randomly selected lipid*omes*, 100 PLS regression models were calculated and cross-validated (Fig. 2.17, a). Necrosis, vital tumor content and stroma together with the tumor entity had the highest numbers of significant models with 45 to 60 out of 100 (Fig. 2.17, a). Certainly, this fraction is not high enough for a diagnostic model, but it identifies these parameters to be mainly influencing tumor lipid*omes* and confirms the results from the pilot study (section 2.2.3, page 53). The Hb content, BMI, gender, and age showed less than 20 significant models (Fig. 2.17, a) and were considered to have less influence on the tumor lipid*omes*.

Lipid panels were identified that were associated to the histopathological phenotype. The results support the findings from the pilot study, as for instance the association of TAGs to necrotic areas (section 2.2.3, page 55). Potentially, the variability within the necrosis content of the tissue samples could explain the highly variable TAG content in the tissue lipid*omes* (Fig.

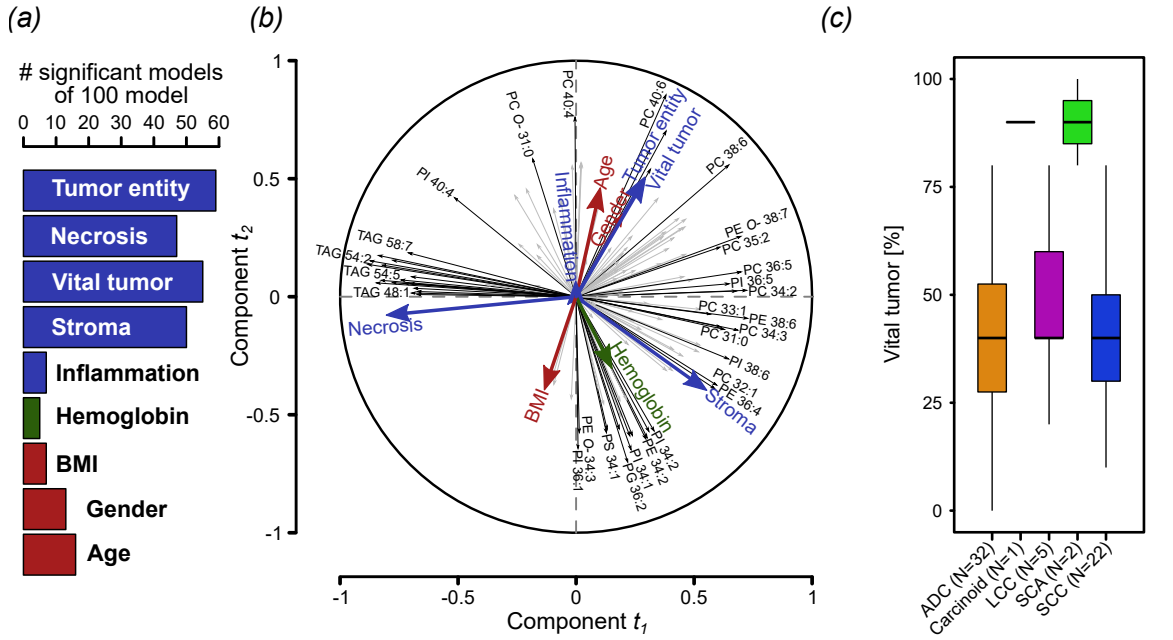


Figure 2.17: **Modeling by partial least-squares regression enabled investigation of lipid*ome* influencing parameters.** *(a)* The barplot shows the number PLS regression models that were significant in cv. *(b)* Correlations of variables with the *t*-components t_1 and t_2 of one exemplary selected PLS regression model. Gray and black vectors represent lipids (predictor variables) and colored vectors represent the analyzed sample parameters (response variables). *(c)* Boxplot giving the distribution of the vital tumor content in ADC (orange), carcinoid (black), LCC (purple), SCA (green) and SCC (blue) tissues. Boxplots were generated with the `boxplot` *R* function.

2.14, d). PE 36:4 and PC 32:1 were associated with the stroma content (Fig. 2.17, b) (Pilot study, Fig. 2.9, page 55). However, the data showed a strong correlation of vital tumor content and the tumor entity on the correlation plot (t_1 and t_2) (Fig. 2.17, b). A possible explanation was revealed by closer inspection of the distribution of the vital tumor content of different tumor entities. The data showed that the vital tumor content was not equally distributed across all tumor entities (Fig. 2.17, c). The NSCLC tissues (ADC, SCC and LCC) had in most cases vital tumor contents between 30 % and 60 %. Overall, the carcinoid and SCA tumors had higher vital tumor contents of 80, 90, and 100 % (Fig. 2.17, c).

For carcinoids, specific lipid*ome* alterations compared to the NSCLC tissues were recognized before in this study (Fig. 2.15, page 55). SCAs tumors were only represented with low numbers in the collected tumor samples and were therefore not compared to NSCLC lipid*omes*. However, specific lipid*ome* alterations compared to the NSCLC tissues are likely. Different tumor entities had different features within their lipid*omes* and the fact that the distribution regarding vital tumor content was not constant across the collected samples potentially caused a bias in

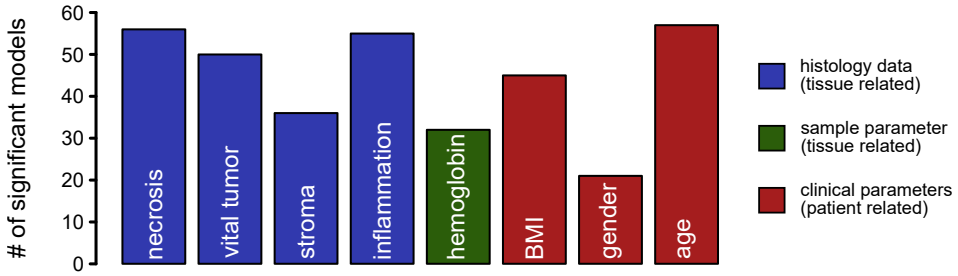


Figure 2.18: **Adenocarcinoma lipid*omes* were influenced by manifold parameters.** Numbers of significant models out of 100 calculated after creation of 100 randomly composed subsets. Significance was determined by the cv procedure. Histopathological data are shown in blue, the sample parameter Hb content in green and clinical parameters in red.

the regression models. Therefore, closer investigation of the lipid*omes* from different histological lung tumor entities was necessary. The obtained information about their specific lipid*ome* influencing properties can be used for more precise data modeling approaches. However, at the actual stage, the number of available samples is still too limited to investigate all these tumor entities. Consequently, a different strategy had to be used for further investigation of tumor lipid*ome* influencing parameters.

Lipid*omes* of adenocarcinoma tissues reflected a complex pattern of clinical and histopathological parameters

To eliminate the influence of the tumor entity on the models, a subset of the original data was used for further modeling. With 32 out of 62 samples, ADC tissues were the largest sub-fraction within the collected tumor tissues. Accordingly, this phenotype was chosen to be investigated with regard to its lipid*ome* influencing parameters. One hundred and twenty four lipids that were present in at least 90 % of all ADC lipid*omes* were included into this analysis.

Overall, the maximal number of significant models was lower than 60 (Fig. 2.18) and did not increase compared to the models calculated with all collected tumor tissues (Fig. 2.17, a). For the histopathological parameters necrosis, vital tumor content and stroma, 56 and 50 models out of 100 were significant in cv. Interestingly, the influence of the inflammation status on the lipid*omes* increased in this analysis (Fig. 2.18) compared to the analysis made before. The BMI and age were significant in 45 and 57 models which supports the hypothesis that patient related parameters influenced tumor tissue lipid*omes*. The gender had a low influence demonstrated by a number of 21 significant models and the Hb content of the tissues correlated significantly in

32 of 100 calculated models. Although Hb was identified to be minor influencing, the amount of residual blood within the tissues might increase the heterogeneity of the tissue lipid*omes* which has to be considered for future studies.

The stability of the models was then examined by closer inspection of the regression coefficients which connected the lipid abundances (predictor variables) with the investigated clinical and histopathological parameters (response variables). A positive regression coefficient indicated that the respective lipid species increased in its abundance when the parameter, the coefficient was calculated for, increased and *vice versa*. A summary of all standardized regression coefficients for the 100 calculated models is shown for the vital tumor content (Fig. 2.19). It is clearly visible from the large whiskers that the variability of all models reached high levels. It was observed that *e. g.* PG species such as PG 34:1, PG 36:2 together with saturated PCs such as PC 32:0 had negative standard coefficients in approximately all models (Fig. 2.19). This means that the abundances of these species decreased with increasing proportions of vital tumor. This observation implies that PG and PC, as components of the pulmonary surfactant, originated in other compartments than vital tumor cells. A number of PC species such as PC 36:1, PC 38:4 and PC 40:6 had positive standard coefficients. In the models, abundances of these lipids increased with larger proportions of vital tumor (Fig. 2.19). These lipids can be associated to cellular membranes which nicely represents the higher cell density within tumor regions compared to necrotic and stroma areas. Therefore, these lipids are interesting candidates as diagnostic markers as well as targets for further investigations to understand the lipid metabolism of tumor cells.

For further interpretation of the results from the PLS regression models, one model was additionally calculated with all 32 ADC lipid*omes* without random split into subsets. The standardized regression coefficients of this model correlated in most cases to the median values of the 100 models and this model is therefore considered to reflect an average PLS regression model (Fig. 2.19) and was used for visualization of further results from the PLS regression analysis (Fig. 2.20). On the factor map of the average model, the individuals distributed according to decreasing vital tumor contents from the top-left to the bottom right corners (Fig. 2.20, b). This further supported the notion that vital tumor content was a major parameter shaping tumor lipid*omes*.

The correlation circle of the average model (Fig. 2.20, b) revealed further information with

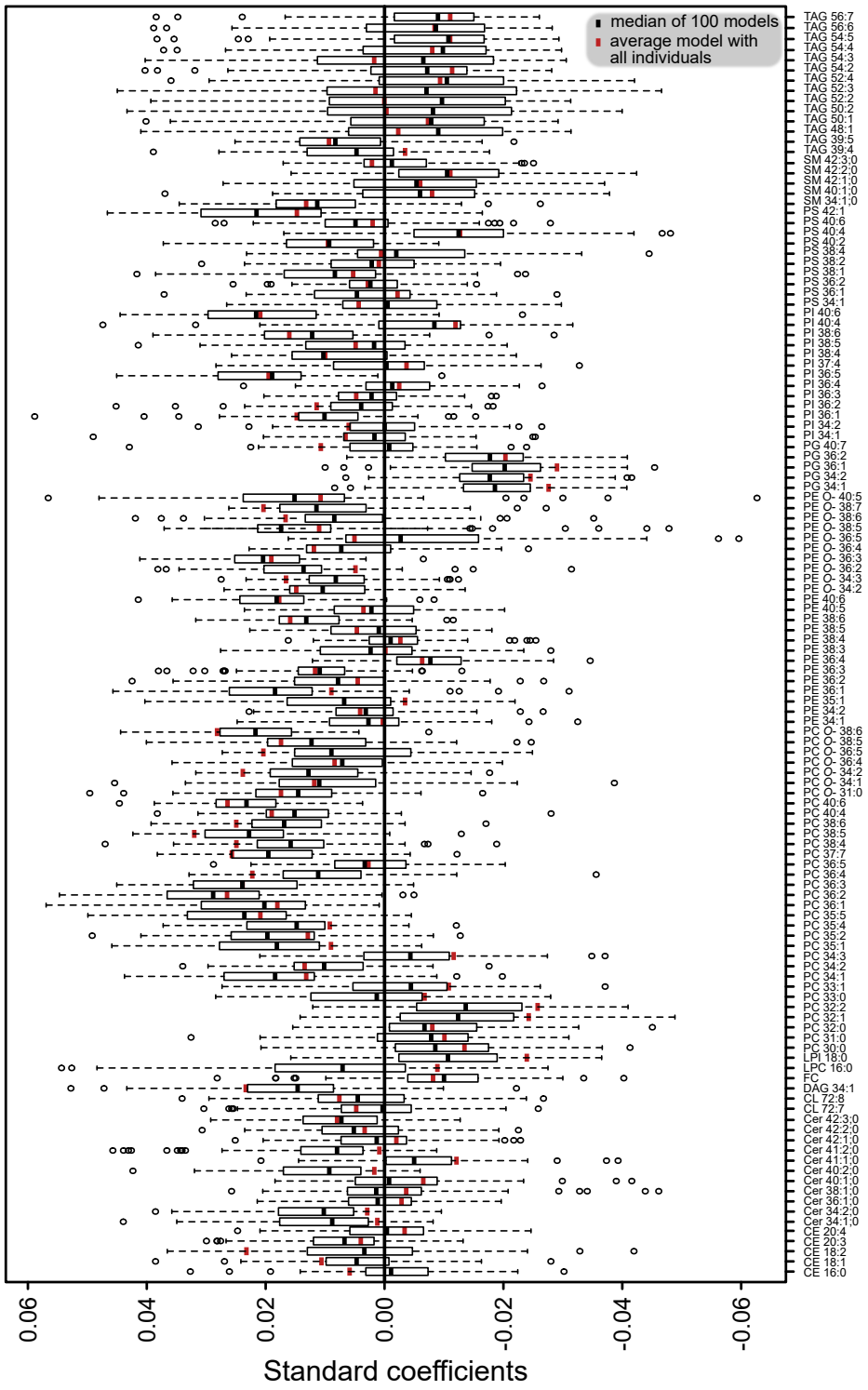


Figure 2.19: **Standardized regression coefficients show specific lipid markers for metabolically active tumor content of adenocarcinoma tissues.** The value of the coefficient reflects the influence of the respective lipid species on the model for vital tumor content. A positive standard coefficient shows a positive correlation and *vice versa*. The boxplots give information about standard coefficients from the 100 calculated PLS regressions with the median in black. Red indicates the standard coefficients of one PLS regression model which was calculated with the complete dataset without creation of subsets and which is assumed to reflect an average model. The boxplots were created with the *R* function `boxplot`.

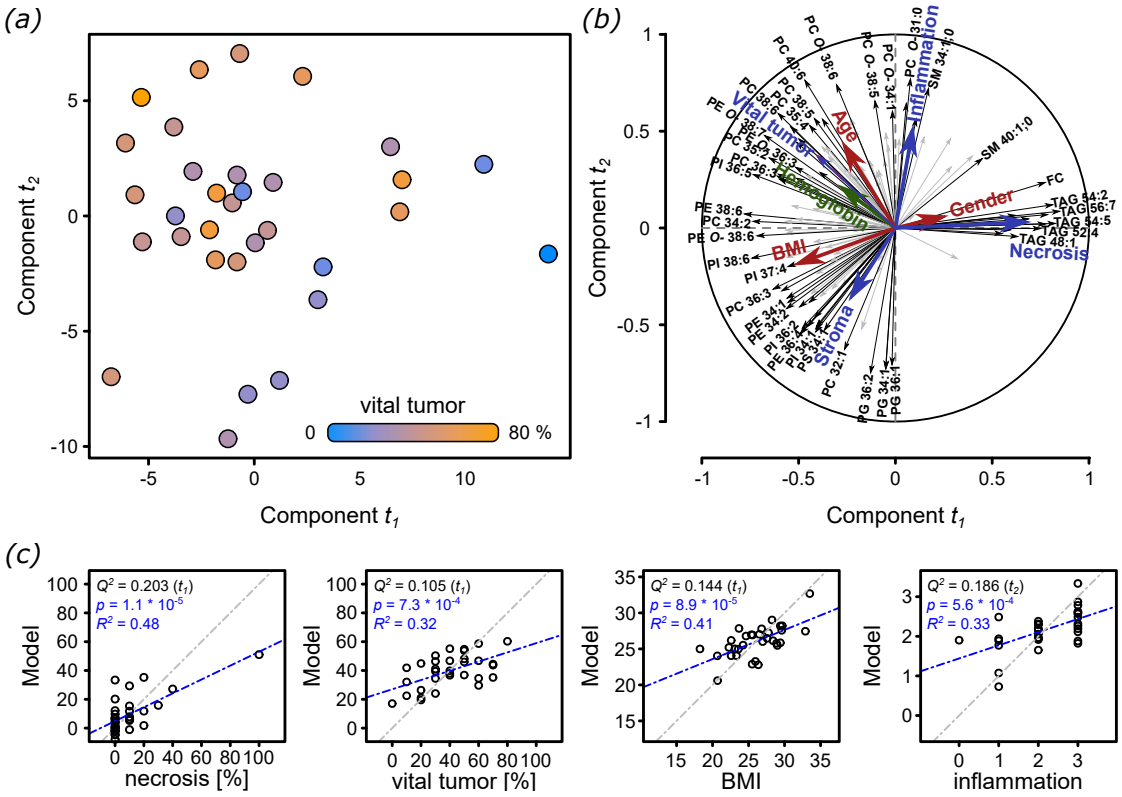


Figure 2.20: **The average partial-least squares regression model showed that adenocarcinoma lipid*omes* correlated to vital tumor content, necrosis, BMI and inflammation.** *(a)* PLS regression factor map showing the individual tumor lipid*omes* that were color coded according to the content of vital tumor with a color scale from blue (low) to yellow (high). *(b)* Correlations of lipid quantities, histology scores and clinical data with the *t*-components. The vectors indicate how strong the variables correlate to the *t*-components and they also show correlations between the lipid*ome* data (black arrows) and the histology scores and clinical data. *(c)* Correlation of original histology scores to scores that were computed from the PLS regression average model for necrosis, vital tumor, BMI and inflammation score. The gray lines give the identity where the ideal correlation is expected. The blue line shows the linear regressions.

regard to specifically altering lipid species. For instance, a correlation of increasing amounts of TAG species with the extent of necrosis was observed and supported the findings obtained in the pilot study (section 2.2.3, page 55). To investigate specific influences of necrotic areas, it would be mandatory to collect samples that are equally covering the whole available range from 0 to 100 % necrosis since highly necrotic samples are underrepresented in the collected samples. PC 32:1 and PE 36:4 together with other anionic GPL species such as PS 34:1 and PG 34:1 correlated positively with the stroma content (Fig. 2.20, b) which negatively correlated with the inflammation status. This indicated a reciprocal relationship of inflammation and stroma. The anionic GPL species, in particular PGs, are known to be involved as immune suppressors in the lung. A loss of these species could support progression of inflammation.

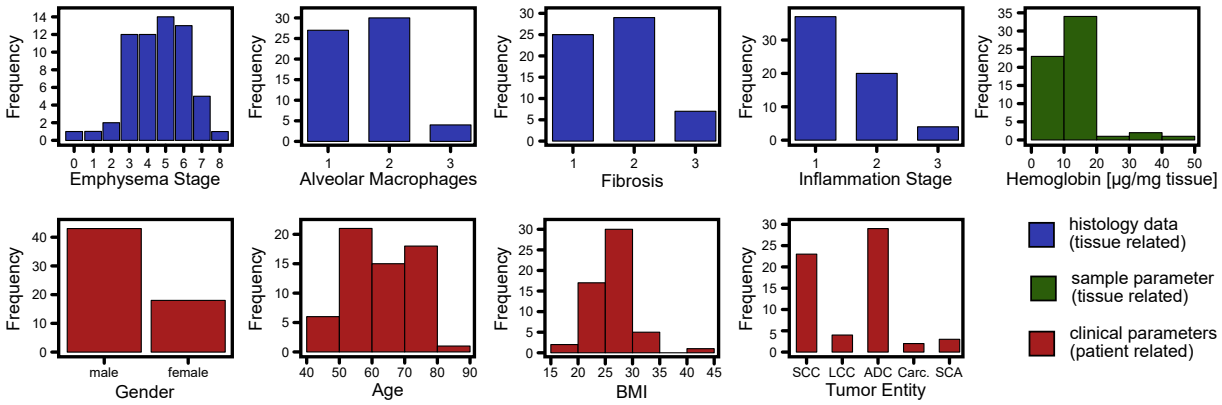


Figure 2.21: **Histological and clinical parameters that were included in partial least-squares regression with alveolar tissue lipid*omes*.** Histograms show the distribution of the histologic parameters (in blue) across all 61 analyzed lung tissue samples. Clinical parameters age, gender, BMI, and tumor entity of the corresponding primary tumor are shown in red. The sample parameter hemoglobin content is displayed in green. Histograms were plotted as frequency distributions with the *R* function `hist`.

The cv for the the average model was significant for necrosis ($Q^2 = 0.203$), vital tumor content ($Q^2 = 0.105$), BMI ($Q^2 = 0.144$) and the inflammation status ($Q^2 = 0.186$) (Fig. 2.20, c). Taken together, these parameters were identified as the main factors shaping the tumor lipid*omes*. However, allover Q^2 values were below 0.2 which indicated that the achieved regression models were not suitable for diagnostic purposes which could be a matter for further discussion. In consecutive analyses, it would be interesting whether similar patterns of lipid*ome* influencing parameters occurred in alveolar tissues.

2.3.5 Tumor entity, aging and gender shaped the lipid*omes* of alveolar lung tissues

Partial least-squares regression analysis revealed lipid metabolic alterations in alveolar lung tissues regarding tumor entity

The pilot study demonstrated that pulmonary emphysema and aging are reflected in the lipid*omes* of the tumor-free alveolar lung tissues. Furthermore, the outcomes from the pilot study indicated that the corresponding tumor entity could also be reflected in the lipid profiles of the alveolar, tumor-free compartments (section 2.2.4, page 57). For further investigation of parameters that characterized lipid profiles of alveolar lung tissues, the same modeling approach was applied to the lipid*omes* of tumor-free alveolar tissues.

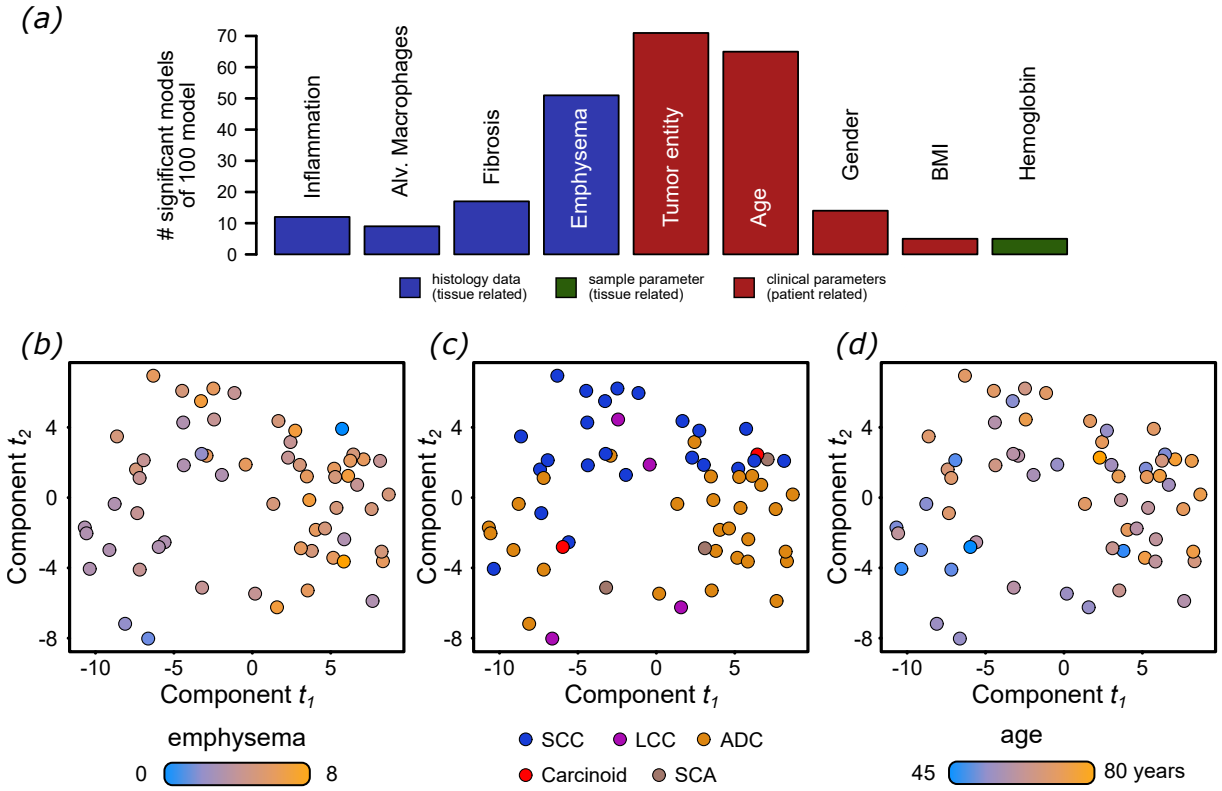


Figure 2.22: **Lipid*omes* of alveolar tissues correlated with clinical data and histopathological scores.** *(a)* The Barplot shows the numbers of significant PLS regression models out of 100 calculated models. The dataset consisting of 61 alveolar tissue lipid*omes* was randomly split 100 times into training and validation subsets. For each subset one model was calculated. Significance was determined by leave-out-one cv with a Q^2 value larger or equal to 0.0975. *(b)* PLS regression factor map (components t_1 and t_2) from a model calculated with all 61 tissue lipid*omes*. The points on the map represent each an individual tissue lipid*ome*. The points are color coded according to their stage of emphysema as noted in the legend. *(c)* Factor map color coded according to the entity of the primary tumor. SCC is marked in blue, LCC in purple, ADC in orange, carcinoid in red and SCA in brown. *(d)* Factor map where individuals are color coded according to the patients age at surgery.

Alveolar tissues were scored for their extent of emphysema as well as for the amount of AM, fibrosis and inflammation (Fig. 2.21). Pulmonary emphysema was observed in stages from 0 to 8 where 51 tissues out of 61 tissues were scored in the stages 3 to 6 (Fig. 2.21). AM and fibrosis, both scored in the stages 1, 2 and 3, were most frequently observed in the stages 1 and 2 with 25 to 30 observations. AM and fibrosis stage 3 scores had only 4 and 7 observations, respectively. Inflammation was most frequently observed in stage 1 (37 observations) followed by the stages 2 (20 observations) and 3 (4 observations). The patient related parameters gender, age, BMI and tumor entity had similar distributions as observed in the tumor tissues (Fig. 2.16, page 69). Hb levels of the alveolar tissues had values up to 50 µg/mg tissue and were higher compared to the tumors. However, the majority of values was lower than 20 µg/mg (Fig. 2.21).

Lipid*omes* from 61 alveolar tissues, including 127 lipid species present in at least 90 % of the samples, were used for PLS regression modeling. As performed for tumors, 100 models were calculated based on randomly selected subsets of the 61 tissue lipid*omes*. Additionally, one average model was calculated including all 61 tissue lipid*omes*. The largest number of significant correlations was calculated for the entity of the corresponding tumor disease (71 models significant out of 100) (Fig. 2.22, a). The fact, that the tumor entity was even reflected in the lipid profiles of tumor-free lung tissues was further indicated on the factor map of the average model where a separation of patients suffering from ADC and SCC was observed along the axis t_2 (Fig. 2.22, c). In addition, 65 out of 100 regression models were significant for aging. This result was also reflected on the factor map where the tissue lipid*omes* sorted from bottom-left to top-right according to the patients' age (Fig. 2.22, d). Pulmonary emphysema revealed significant correlations in 51 out of 100 regression models (Fig. 2.22, a). The factor map indicated a distribution of the tissue lipid*omes* with increasing emphysema stages along axis t_1 (Fig. 2.22, b). Both parameters, aging and emphysema, were associated to t_1 which further fits into the observation made in the pilot study regarding the interplay of emphysema progression and aging (section 2.2.4, page 56). For parameters such as inflammation stage, AM, fibrosis, gender, BMI and the Hb content, less than 20 significant models and were computed and these factors were assumed to be minor influencing.

Taken together, the correlation of emphysema with aging as observed in the pilot study was confirmed here. However, these trends were weak compared to the impact of the tumor entity on the lipid profiles. In a consecutive step, a statistical test could provide information about the predictive power of the alveolar lung lipid*ome* to diagnose the tumor entity in cancer patients.

Tumor-free alveolar lung tissues of adenocarcinoma and squamous-cell carcinoma patients are distinguished in their lipid profiles

To investigate whether the lipid*ome* provides diagnostic markers that differentiate between ADC and SCC patients, the lipid profiles of alveolar tumor-free tissues from them were compared using the U test (Fig. 2.23, a). A panel of lipids was identified which was significantly changed between these two groups (Fig. 2.23, a). PC species such as PC 34:2, PC 31:0, PC 38:6 and PC *O*- 36:5 were systematically increased in the lipid*omes* of tumor-free lung tissues of ADC patients (Fig. 2.23, a, d and e). The majority of the systematically altered PC species was low

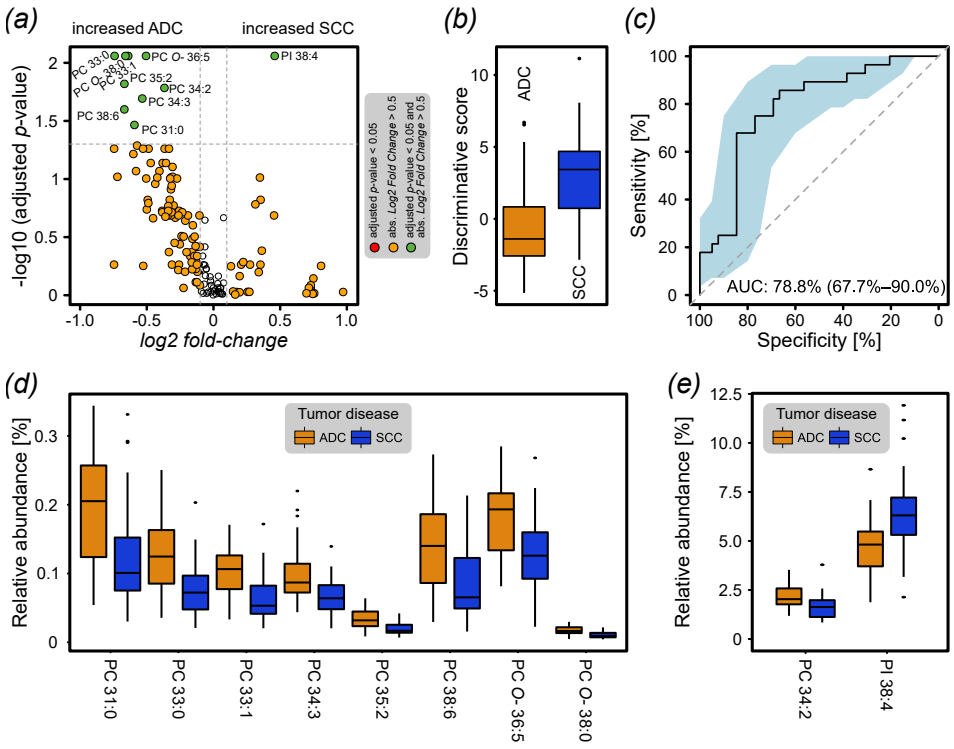


Figure 2.23: **Wilcoxon Mann-Whitney U test revealed distinct differences between lipid*omes* of tumor-free alveolar lung tissues of adenocarcinoma and squamous-cell carcinoma patients.** *(a)* Volcano plot visualizing *p*-values (U, unpaired, two-sided), and log2 fold-changes comparing alveolar tissue lipid*omes* of ADC patients (N = 39) with alveolar lipid*omes* from SCC patients (N = 28). Each point represents a lipid species. The color code corresponds to their adjusted *p*-value (Benjamini-Hochberg FDR) and log2 fold-change. *(b)* Boxplot that shows the score to discriminate alveolar lipid*omes* from ADC and SCC patients. The score is based on the green marked lipid in panel a. The score was calculated in the same manner as noted for figure 2.14. *(c)* ROC curve based on the discriminative score. The blue area represents the 90 % confidence area. *(d)* Boxplot representing selected lipids (except PC 34:2 and PI 38:4) from panel in a concentration range up to 0.35 % relative abundance. *(e)* Boxplot representing lipids PC 34:2 and PI 38:4 in a concentration range up to 12.5 % relative abundance. Boxplots were drawn using the *R* function `boxplot`.

concentrated lipids with relative abundances lower than 0.3 %. PI 38:4 was increased in alveolar lipid*omes* of SCC patients. A number of lipid species with odd numbered FAs was significantly changed between ADC and SCC patients (Fig. 2.23, d).

The calculated discriminative score based on the selected diagnostic panel (Fig. 2.23, b) enabled good discrimination of patients with ADC and SCC diagnosis (AUC 78.8 %, Fig. 2.23, c). Due to their diagnostic potential, the selected lipids could be suggested as candidates for specific tumor markers in the lung. Nevertheless, it has to be considered that healthy controls were missing in this analysis since healthy lung tissues are not accessible from lung cancer patients and tissue samples from healthy donors are rare. The capability of this approach should

be proven by incorporation of healthy controls. As already mentioned in the pilot study, the discovered lipid*ome* alterations in lung cancer might be a limitation of the utilization of residual tissue material from lung cancer patients, especially due to the unknown impact of smoking on the lung lipid*ome* composition and its interplay in carcinogenic processes (section 2.2.4, page 57). For future analysis, it would be necessary to include the smoking history of the patients to evaluate possible connections of cigarette smoke consumption and lipid metabolic processes. In a consecutive step, a subset of the collected cohort that exclusively comprises lipid*omes* from patients suffering from the same tumor disease is analyzed.

Emphysema, age, and gender mainly influenced alveolar lipid*omes* of ADC patients

The previous section showed alterations between the lipid*omes* of the tumor-free alveolar lung tissues of ADC patients and SCC patients. To learn more about the lipid*ome* composition of human alveolar lung tissues, independent from the cancer disease, the subset consisting of the alveolar tissue lipid*omes* was further analyzed. Accordingly, lipid*omes* including 127 lipid species present in 90 % of the samples that were provided from 29 ADC patients were used as input data for PLS regression modeling. The largest number of significant PLS regression models was achieved for the emphysema stage with 83 significant models out of 100 calculated models (Fig. 2.24, a). Aging and gender parameters had about 80 models significant out of 100. Inflammation, AM and fibrosis had 66, 44, and 62 significant models. Moreover, BMI and Hb content had 13 and 23 significant models and were identified as minor influencing parameters on the lipid*omes* of alveolar lung tissues.

To evaluate which parameter had the greatest impact on the lipd*ome* composition, the Q^2 values that were obtained in the cv procedure were inspected. Emphysema and age reached high Q^2 values along component t_1 in most cases (Fig. 2.24, b). From this observation it might be concluded that aging and emphysema share features influencing the lipid*omes* of alveolar tissues which further supports the findings from the pilot study (section 2.2.4, page 56, [146]). A gender-specific effect on the alveolar lipid*ome* composition was demonstrated by a high number of significant regression models for gender along component t_2 (Fig. 2.24, c). Along component t_3 none of the investigated parameters reached significant Q^2 values (Fig. 2.24, d). Taken together, the three parameters pulmonary emphysema, aging and gender were identified with high numbers of significant models and high levels of significance.

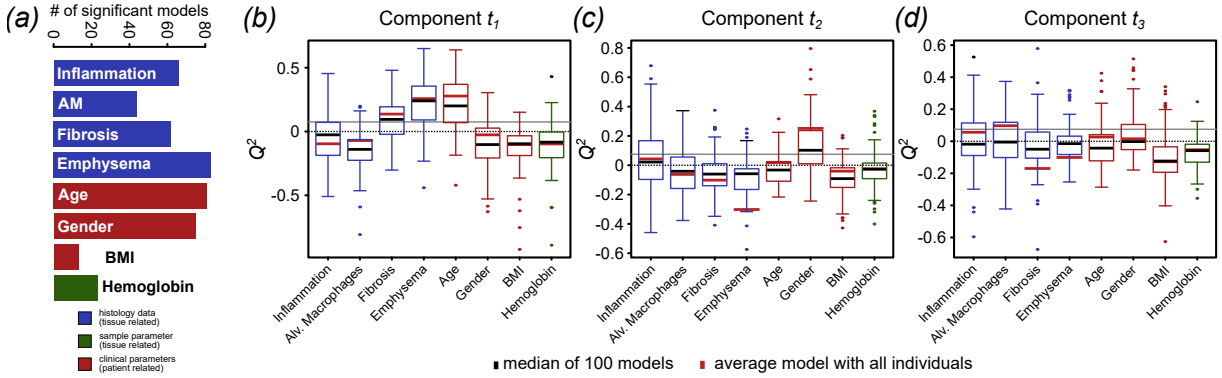


Figure 2.24: **Partial least-squares regression analysis of tumor-free alveolar tissue lipid*omes* of adenocarcinoma patients.** Alveolar tissue lipid*omes* of 29 ADC patients were included into the analysis. *(a)* Barplot showing the numbers of significant PLS regression models out 100. The dataset consisting of 29 lipid*omes* was randomly split 100 times. For each subset one model was calculated. Significance was determined by cv with a Q^2 value greater or equal to 0.0975. *(b)* Q^2 values for component t_1 and the predictor variables. Boxplots give the variance of Q^2 values for 100 models. Boxplots were drawn with the `boxplot` function from *R*. The black line shows the median Q^2 values of 100 models. The red line indicates the Q^2 of one model calculated with all 29 lipid*omes* and without subset-creation of the dataset (average model). *(c)* Q^2 values on component t_2. *(d)* Q^2 values on component t_3.

These observations were specified on the PLS regression factor map. The tissue lipid*omes* were sorted according to increasing values of the emphysema stage and age with increasing values of t_1 (Fig. 2.25, a and b). Along axis t_2, a separation of tissue lipid*omes* from male and female patients was observed (Fig. 2.25, c). The originally determined values of the parameters emphysema, age and gender correlated positively with the predicted values from the PLS regression models. This showed that it could be possible to put these parameters into a predictive model together with lipid*ome* data (Fig. 2.25, d-f).

Next, it was of interest how the histopathological and clinical parameters correlated with specific lipid species. This information is available from the variable correlations (Fig. 2.26) and from the regression coefficients (Fig. 2.27). Both, emphysema and aging positively correlated with component t_1 (Fig. 2.26, a and b) which was completely consistent with the results from the cv procedure and the observation on the factor map. PG species such as PG 36:2, PG 34:1 and PG 38:5 had negative correlations with emphysema stage and aging (Fig. 2.26, a and b). Additionally, these species had together with shorter chain PI species such as PI 34:1, PI 36:2 and PI 36:4 negative standardized regression coefficients for emphysema and age (Fig 2.27, a and b). This indicates that the relative proportions in the lipid*ome* composition of alveolar lung tissues regarding PG and PI decreased with increasing emphysema grade and/or age. In-

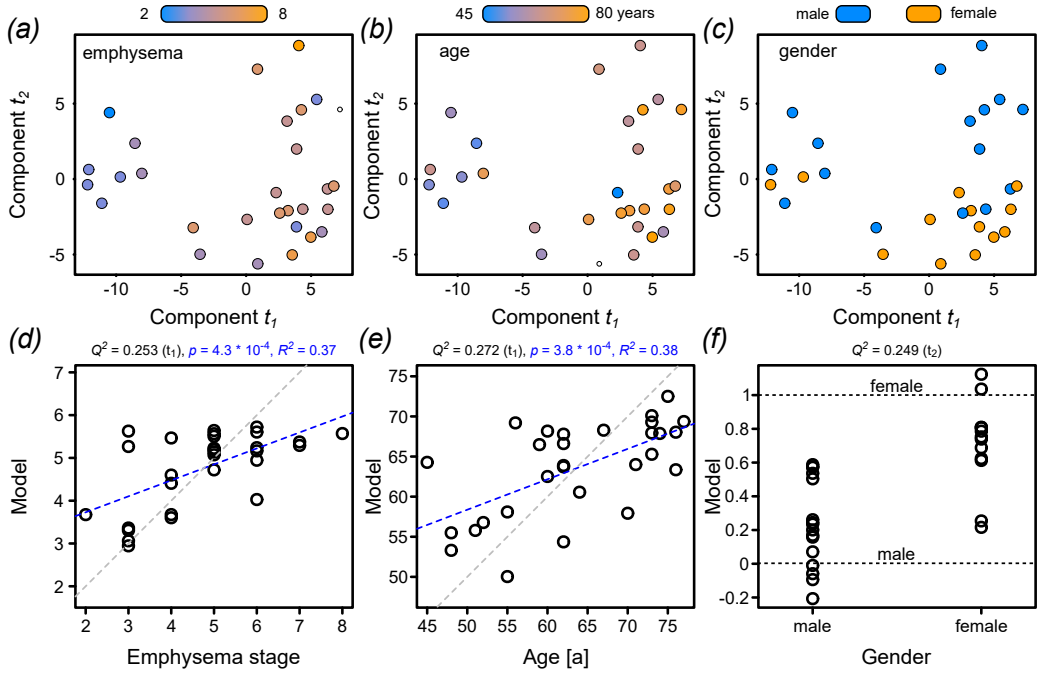


Figure 2.25: **Emphysema, aging and gender are most influencing parameters for alveolar lung tissue lipid*omes*.** *(a)* PLS regression factor map of the average model showing components t_1 and t_2. Each dot represents a tissue lipid*ome* and is color coded according to its emphysema stage. *(b)* The same factor map with color coding according to age. *(c)* The same factor map with color coding according to gender. *(d)* Linear correlation of emphysema stage and predicted values by the average model. *(e)* Correlations for age. *(f)* Correlations for gender. The linear regressions are shown in blue. The gray line indicates the ideal model. Q^2, p and R^2 values are noted above the panels.

terestingly, PG 36:0 was observed as the only PG species with positive standard coefficients (Fig. 2.27, a and b). Additionally, a number of Cer species such as Cer 34:1;0, Cer 38:1;0 and Cer 40:2;0 had positive standard coefficients for emphysema and age, indicating an increase of those species with increasing values for emphysema and age. Moreover, there were PC species including PC 34:1 and PC 36:4 which were increased with higher emphysema stage and age.

Due to the negative correlation of gender and component t_2, gender showed an approximately orthogonal behavior compared to emphysema and aging as shown in the correlations (Fig. 2.26 a and b). The distribution pattern of standard coefficients for gender was distinct from emphysema and age (Fig. 2.27). This demonstrated that the lipid*ome* features altered with gender were completely distinct from the lipid*ome* features which were relevant for emphysema and aging. Many of the 127 lipid species that were included into this analysis had negative standard coefficients. In the input data, the categories for male and female patients were expressed with the numerical factors 0 and 1 which means that lipids with positive standard coefficients were

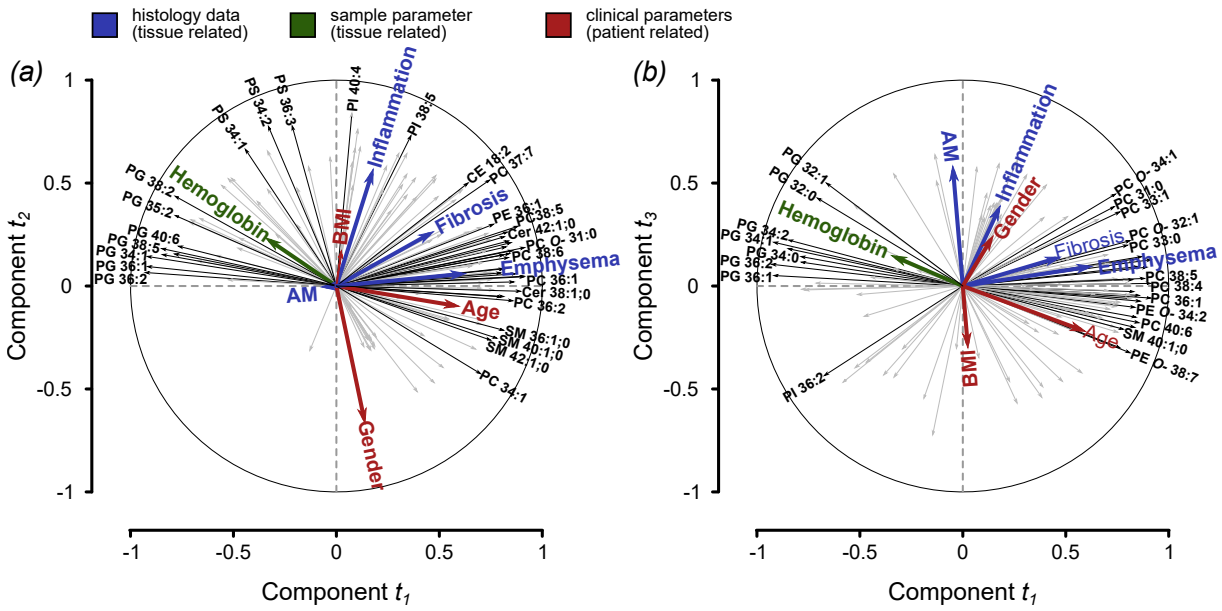


Figure 2.26: **Specific correlation of lipid*omes* from alveolar lung tissues of adenocarcinoma patients to histopathological and clinical parameters.** Results from the same analysis as shown in figure 2.24. *(a)* Correlation of lipid quantities, histology scores and clinical data to the *t*-components t_1 and t_2. The vectors indicate how strong the variables correlate to the *t*-components and they also show correlations between the lipid*ome* data (black arrows), the histology scores and clinical data. *(b)* Correlation for *t*-components t_1 and t_3.

increased in females and *vice versa*. In males, the models showed systematic increased levels of CE species such as CE 18:2, of PE species such as PE 34:2 and PE 36:4, of long-chain PI species such as PI 40:6, and of PS species such as PS 36:4 and PS 40:6 (Fig. 2.27, c) whereas female patients showed increased levels of PC 31:0, PC 32:0 and PC 34:1 as well as a number of long-chain SM species together with TAG species such as SM 40:2;0, SM 42:2;0, TAG 50:2 and TAG 54:3.

A complex pattern of lipid*ome* influencing parameters was present in alveolar human lung tissues. The interplay of emphysema progression and aging was identified to play a key role in the pulmonary lipid metabolism. This finding supported earlier reports on the interplay of aging and progression of emphysema [149]. Furthermore, the models revealed a gender influence on the lipid*ome* composition of alveolar lung tissues. In future, more parameters should be included to get a more comprehensive picture of parameters altering the lipid*ome* composition. In this regard, the effect of patients' lifestyle with respect to *e. g.* nutrition and smoking should be further investigated. The presented modeling strategy revealed promising candidates for further investigations and to understand the molecular processes of progression of the pulmonary emphysema.

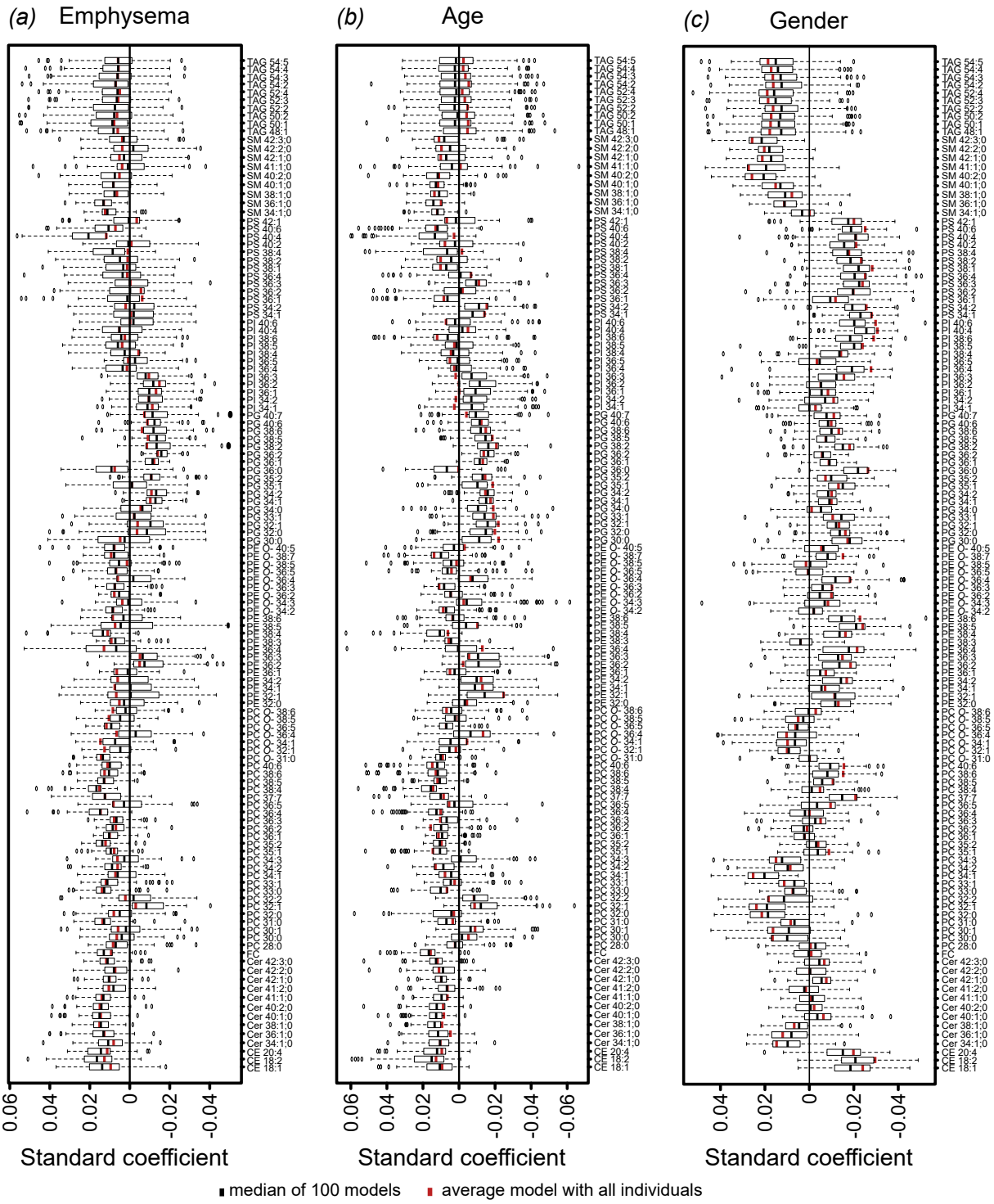


Figure 2.27: **Regression coefficients reveal specific lipid*ome* features for pulmonary emphysema, aging and gender.** *(a)* Standardized regression coefficients for emphysema from the 100 calculated PLS regression models. The value of the coefficient reflects the influence on the model. Boxplots were visualized with the `boxplot` function from *R*. The boxplots give information about the stability of standard coefficients from the 100 calculated PLS regressions with the median value indicated by a black line. The red line shows standard coefficients of one PLS regression model calculated with the complete dataset without creation of subsets. *(b)* Standard coefficients for age. *(c)* Standards coefficients for gender.

3 Discussion

3.1 The analytical strategy allowed a thorough investigation of human lung lipid*omes*

The presented study provides a comprehensive description of a lipid metabolic snapshot for alveolar lung tissue samples and corresponding tumor tissues. The analytical strategy based on shotgun lipid*omics* screens delivered high quality data and allowed a thorough investigation of lipid profiles from human lung tissues. Furthermore, the analytical strategy involved carefully coordinated ionization, acquisition and quantification steps, which are discussed in the following paragraphs.

The ionization strategy utilizing ammonium chloride as additive enabled unambiguous lipid identification. The analytical approach was customized to enable clear identification and quantification of the anionic GPLs as well as PCs in the negative ion mode by the accurately determined *m/z*. Accordingly, the ionization strategy was adapted by utilizing NH_4Cl as additive in the sample matrix to force the lipids from the classes of PC, DAG and Cer to form $[M+Cl]^-$ adduct species in the negative ion mode. This modification enhanced the analytical sensitivity and prevented signal interference of PC and PS.

When utilizing NH_4Cl, it needs to be considered that chlorine has two main abundant natural occurring isotopes, ^{35}Cl and ^{37}Cl. ^{35}Cl has about 76 % natural abundance and ^{37}Cl about 34 %. As consequence, there is an isotopic split into ion species consisting ^{35}Cl and ^{37}Cl with a ratio of 3-to-1 in their signal intensities. Consequently, additives that do not show this isotopic splitting but still give unambiguous signals could be a matter of future studies.

The mass spectrometric strategy provided lipid profiles that represented a comprehensive lipid metabolic snapshot of lung tissue samples. The utilized lipid*omics* screening strategy yielded in stable lipid profiles with sufficient precision to describe the biological variability within the analyzed cohort. This acquisition strategy recorded all available mass spectrometric features expressed by an accurately determined *m/z*. Afterwards, specific signals were assigned to lipid classes and species that were known to be present in humans. Therefore, not the full spectral information was used for data interpretation. This filter removed noise and background signals from the data which drastically reduced data size. Utilizing this acquisition strategy, the presented workflow differs from previously performed studies on human lung tissues. Manifold of those studies were focused on biomarker discovery lacking sufficient identification of the targeted molecules. These studies, for instance performed by Lee *et al.* [74] and by Guo *et al.* [4], investigated all available mass spectrometric features, expressed by an *m/z* value, for their diagnostic potential. Then, the full spectral information was used for data analysis. From these up to thousands of features, only a few diagnostic markers were selected and assigned to lipid species identifications. This approach, widely applied in the field of metabol*omics*, has the advantage to be untargeted and to be without any pre-assumptions. The disadvantage is that a complete picture of the lipid metabolic condition is missing which is essential to understand underlying metabolic processes *e. g.* in disease progression. Furthermore, the study by Lee *et al.* [74] investigated lipid*ome* alterations between the histologic phenotype by MALDI-MSI which is a powerful tool for localization of specific molecules but lacks quantitative information. Two other studies were performed by Marien *et al.* [75, 84]. With quantitative phospholipidome profiling they described human lung tissues and tumors but did not cover the important PGs and PAs due to their analytical approach based on QqQ MS.

This dissertation applied a screening strategy that exclusively utilized mass spectrometric signals that were assigned to a known lipid species. Due to this comprehensive identification and assignment strategy the outcomes might be correlated to metabolic pathways to draw functional associations. In contrast to the untargeted approach, the utilized screening strategy is not capable to discover new compounds involved in metabolic processes since the lipid*omics* screen approach is directed exclusively to known lipid compounds. Furthermore, the MS-technique alone is not suitable for *de-novo* determination of molecular structures. For such purposes, other techniques such as crystallography and nuclear magnetic resonance spectroscopy have to

be used in conjunction with mass spectrometric fragmentation patterns. However, as soon as a new compound is discovered which could potentially play a role in lung diseases, it is possible to search the datasets already obtained from this study with regard to their full spectral information for this new compound without having to carry out new cost and time intensive laboratory experiments.

In continuation of this project, the data acquisition strategy for shotgun lipid*omics* screens can be improved in regard of more confident and accurate lipid species assignments and higher sensitivity yielding in a higher lipid*ome* coverage. The association of high resolution survey scans with specific MS^2 experiments combines advantages of *top-down* and *bottom-up* approaches. Quantifier ions are then selected from high resolution MS^2 spectra which increases the specificity for lipid identification and the sensitivity for lipid detection. For further improvement, the full *m/z* range of the survey scans from *m/z* 350 to 1000 can be sequentially scanned in small windows of 150 Da as proposed before [150]. Afterwards, a survey mass spectrum covering the whole mass range is generated by stitching the individual 150 Da mass window acquisitions. This acquisition strategy increases the sensitivity for detection which results in a more stable quantification. Incorporation of MS^2 data further allows identification of FAs connected to the respective lipid species which provides valuable information which is necessary for instance to calculate LUX scores and has become a standard procedure in many lipid*omics* laboratories.

Limits of the quantification strategy. In this study, ISDs containing FA 17:0 were selected. Quantification of PE and PC was based on a diether (PC *OO*-/PE *OO*-) reference compounds.

Diether lipids do not naturally occur in human which makes them useful standards for precise quantification. However, later experiments showed that small amounts of lipids with odd numbered FAs might be present in human lung tissues. The signal intensities of these natural occurring odd numbered FAs were low compared to the amount of ISDs added, which allowed useful data to be obtained with FA 17:0 based quantification. However, the analytical precision was reduced utilizing these standards. The diether lipids PC *OO*- and PE *OO*- were not present in human lung lipid*omes*. Nevertheless, it was observed that ionization efficiencies were unequal for ester- and ether lipids. This observation was made by comparing quantification results from the positive and negative ion mode where the quantities for ether lipids (PC *O*-/PE *O*-) were comparable in both ion modes and a systematic offset was observed

for the lipids with ester-linked FAs (PC/PE). This systematic error reduced the analytical accuracy for lipid quantification. Furthermore, *ether* linked side chains do not undergo the same fragmentation mechanisms like *ester* linked FAs. Accordingly, diether lipids are not suitable as ISDs for quantification on MS^2 level. The third limitation of the utilized ISD substances was that they did not cover all targeted lipid classes such as Cer, DAG and PG. This was done to reduce cost for expensive analytical standards. As consequence, quantification of these lipid classes had to be referred to ISDs from other classes which reduced the analytical accuracy of quantification. Therefore, it was not possible to report absolute molar quantitative results in this study.

To overcome these limitations, an improved quantification strategy could reveal higher diagnostic precision. In follow-up experiments, an optimized set of ISDs might be used. Today, standard mix solutions for lipid*omics* are commercially available that contain isotope labeled substances from GPLs, GLs, cholesterols and major abundant SL classes. Due to the isotope label, these ISDs provide mass spectrometric signals that do not interfere with compounds present in the sample matrix which enhances the analytical precision. The isotope labeled ISDs are chemical identical to the target compounds and undergo therefore same ionization and fragmentation mechanisms which enhances analytical accuracy. The fact that ISD solutions are readily prepared commercially available reduces potential errors during standard preparation and reduces cost. In this work, quantification was achieved by direct referencing of ISD intensity and analyte intensities. This could be improved by introducing a calibration curve for quantification to obtain a more realistic relationship between signal intensities and quantities. First experiments in this regard were performed and confirmed linearity of about three orders of magnitude. Other analytical parameters including limit of detection and limit of quantification can be determined with this approach. Improved quantification strategies will be suitable to report absolute molar quantities which enables metabolic investigations on cellular level [151]. However, large scale lipid*ome* analyses for clinical applications will remain challenging. Ideally, each compound that is targeted within the screening protocol has to be referenced to its own standard. In case of a screening which is targeting 300 lipid species, also 300 standard substances would be required, which is practically not feasible. Therefore, in such screening approaches still compromises in regard to the selection of the standards have to be made due to the fact that hundreds of analytes should be quantified in one analytical run. An intersting

approach was recently proposed by Rampler *et. al* [152] who used ^{13}C labeled lipid*omes* from yeast for quantification in lipid*omics* screenings. However, this approach heavily enhances the complexity of the acquired mass spectra which could reduce the analytical sensitivity.

3.2 Lipid profiles from lung tissues and tumors opened new perspectives in pathogenesis

This study showed marked differences between the lipid profiles of alveolar lung tissues and corresponding tumors. Furthermore, specific lung pathologies as well as histopathological phenotypes correlated with the lipid profiles of the analyzed tissue samples. The way in which lung tissue lipid*omes* reflect specific events in pulmonary pathologies is discussed below.

The lipid*omes* of alveolar and tumor tissues are distinguished by their ratio of surfactant and membrane lipids. This study showed a general shift in the lipid profiles of alveolar lung tissues and lung tumors. In particular, the ratios between the lipids assigned to pulmonary surfactant, such as PG 34:1, to the membrane-associated lipids including PC 34:1 allowed clear discrimination of alveolar and tumor tissues. Those lipids were suitable as diagnostic markers for tissue identification. Accordingly, this dissertation confirmed that PG molecules might play an important role in the pulmonary lipid metabolism, but this was little surprise due to their important function as components of the pulmonary surfactant. Therefore, alveolar lung tissues marked a characteristic fingerprint of surfactant-associated lipids which was less pronounced in the corresponding NSCLC tissues. Interestingly, SPs (*e. g.* SP-A) were identified to be present in ADCs and are used in routine diagnostics of them [153, 154]. For future investigations it would be of high interest to associate abundances of surfactant lipids with SP quantities to evaluate whether their ratios change between alveolar lung tissues and ADCs. Tumor tissues, exhibited increased levels of cell membrane-associated lipids which nicely reflects the higher cell density in tumors compared to alveolar tissues. Overall, tumors had higher levels of TAGs which is a form of energy storage and could be connected to increased energy consumption. All these observations confirmed the unique properties of the tumor lipid*ome* which is completely distinct from alveolar lung tissues.

The limitations of this study remain in the heterogeneous structure of the tissue samples themselves. Due to its bronchial structure the lung is not homogeneously exposed to pathogens which can lead to pathological conditions such as the emphysema being partially differently developed. Therefore, the examined samples demonstrated only a small section of the lung. The tumors also had a heterogeneous composition which was clearly indicated by their different extent of necrosis and stroma. When comparing lipid*omes* from these heterogeneous tissues, the findings might be limited to the major changes while the smallest lipid metabolic alterations might conceal in the tissue heterogeneity. Especially, this was the reason why bronchial tissues were not included into this study. The most interesting compartment of a bronchus for investigation of lung pathologies, such as COPD, is the bronchial epithelium which is unfortunately only a small fraction of the whole bronchus. Potentially occurring lipid metabolic alterations during pathogenesis in the bronchial epithelium would then be superimposed by the dominant portions of muscles and connective tissues in the bronchus.

As alternative sampling approaches, fractionation techniques for tissue samples are available. Another study mechnically isolated bronchial epithelia using sterile swabs and showed significant differences in the lipid profiles of these isolated epithelia compared to the whole bronchus [155]. The advantage of this approach is that it could be applied without any media, cell culturing and detergents which could modify the lipid profiles. However, lipid*omes* were only analytically accessible by LC-MS analysis due to polymer contamination from the swab which interfered with the lipid signals in shotgun lipid*omics* screens. Unfortunately, this methodology is not easily applied to remove alveolar epithelia due to their different anatomical structure. Another localization-resolving technique is laser capture microdissection which has been applied in lipid*omics* research before [156] and is suitable to isolate specific fractions from tissue slices. Tissue cryo-sections can be selectively ablated by picosecond infra-red laser pulses without chemical modification of the containing molecules as it was shown for complex proteins before [157] while lipids are under investigation (unpublished result). However, the penetration depth of the laser is about 100 to 200 µm and therefore too high for selective analysis of epithelia rather than single cell populations. However, it reveals a promising approach for future studies. All these applications have a high potential to focus the analyses to tissue substructures and to selectively investigate pathological events such as carcinogenesis, chronic inflammatory processes and granulomas with a quantitative lipid*omics* approach.

A potential link between tumor malignancy and lipid metabolic processes. The lipid*omes* of different tumor entities had distinct features. A diagnostic score was calculated from the lipid*ome* data to differentiate between ADC and SCC tissues. Cers clearly separated the low malignant carcinoid tumors from NSCLC tissues. A significant decrease of surfacant-associated lipids including PGs and saturated PCs was detected in the carcinoid lipid*omes* compared to the NSCLCs. Recently, tumor malignancy was connceted to the presence of the FA receptor CD36 together with highly saturated FAs [158]. Pulmonary surfactant-associated lipids are known to contain high proportions of these saturated FAs [159] and it was shown before that CD36 mediated their cellular uptake [160]. Potentially, there might be a connection between lung cancer malignancy and the presence of pulmonary surfactant. This could reveal new strategies for therapeutic targets. With these observations, this dissertation demonstrated potential indications that tumor malignancy and the lipid metabolism could be closely connected parameters.

Nevertheless, this observation needs to be treated with caution since the numbers of cases, especially for carcinoid tumors, were relatively low. The tissue heterogeneity could also influence the results as already discussed. The hypothesis about a potential connection of tumor malignancy and lipid metabolic processes could not be proven from the data available from this dissertation since no information regarding tumor staging and malignancy were included.

Due to the heterogeneity of histopathological and clinical parameters, it will be still necessary to collect more samples which makes it possible to selectively analyze sub-cohorts sharing the same tumor entity, age or gender. The clinical data could be completed by considering information about the stage of the respective tumor. The so-called TNM staging system is routinely used in diagnostics [161] and provides information on the size of the primary tumor (T), whether lymph-nodes (N) were infiltrated by tumor cells and whether the tumor had formed metastases (M). Combining this information with the lipid profiles of the tumors might show connections between lipid metabolic processes and tumor progression.

Lipid profiles of tumor tissues reflect a complex pattern of lipid*ome* influencing parameters. This study revealed a complex pattern of features that were influencing the lipid*omes* of tumor tissues beyond the tumor entity. Most strongly, lipid*omes* of tumor tissues showed that the tissue morphology is reflected in the lipid*omes*. Regression analysis indicated specific lipids correlating to the proportion of vital tumor, necrosis and stroma. For instance, the data

showed that TAG correlated with necrotic regions within the tumor samples. Further attributes including inflammation status and BMI were identified to be tumor lipid*ome* influencing. This demonstrates that general patient related factors are reflected in the tissue lipid*omes* which is important in developement of personalized medicine concepts [162].

In this study, attempts were made to include as many parameters as possible in the analysis in order to examine them with regard to their effects on the lipid profiles. However, a fairly complex pattern appeared and it is unclear whether there were unknown influencing features. Furthermore, not all included factors were independent from each other, which was indicated for instance by the observation that the tumor entity and the tumor content correlated. These co-correlations made it difficult to associate specific lipid metabolic features to the attributes that were influencing the lipid*omes*. A further limitation is that extreme values of the parameters were often underrepresented. This was most strongly observed for the necrosis content where only two tissue samples had contents above 50 %.

For future studies, the sampling strategy can be revised. Depending on availability, several samples can be taken from the same removed lung tumor for a specific analysis of the tissue heterogeneity. Then, all patient related features would be constant that the analysis could be focused on the tissue composition. This information can then be used to create a model for further investigation of the influences of patients lifestyle parameters on the tissue lipid*omes*.

Carcinogenesis induces lipid metabolic alterations in tumor-free alveolar lung tissues. This study revealed specific alterations within the lipid profiles between tumor-free alveolar tissues from ADC and SCC patients. This finding indicates a general lipid metabolic reorganization occurring in lung cancer. However, the main abundant lipid species did not alter in their determined quantities between both groups most likely due to the fact that these tissues still maintained their normal functions. Moreover, a set of lower abundant PC species was identified with alterations in their abundances between both groups.

The study design allowed it only to compare the alveolar tissue lipid*omes* between cancer patients. However, the study delivers no information about the lipid metabolic status in healthy human lung tissues.

Thus, the findings obtained in this dissertation have to be validated in a larger cohort and incorporated with other tumor entities. Furthermore, control samples of healthy lung tissues

would be mandatory to include for properly investigation of the metabolic reorganization in alveolar lung tissues in carcinogenesis. However, healthy lung tissue samples are not easily accessible since these tissues from healthy donors are rare. It would be of high interest whether this set of lipid species is also significantly changed in their abundances in easier accessible sample types such as BALF. The discovery of specific disease markers would open possibilities in development of early stage diagnostic tools for lung cancer which are urgently needed.

Specific lipid metabolic signatures were associated to pathogenic events in alveolar tissues. This study identified a complex pattern of parameters influencing the lipid*omes* of alveolar lung tissues. The highest influences were identified to be induced by aging and pulmonary emphysema. Today, the basic mechanisms of emphysema progression are not completely known although there are some studies connecting emphysema to alterations in the Cer metabolism as well as to oxidative stress in animal models [64, 163, 164]. These observations were not confirmed here. However, other pathological events such as inflammation and fibrosis were further identified to have specific influences on the lipid metabolic status of lung tissues. Additionally, the study revealed lipid*ome* alterations between the genders.

This high complexity makes it challenging to identify specific lipid metabolic signatures for one pathological event. Furthermore, the data indicated a possible connection of emphysema progression, aging. It was hypothesized before that emphysema progression and COPD are diseases of accelerated lung aging [149] which closely connects these parameters. This interconnection makes it impossible to extract specific lipid metabolic features for only one of these parameters. Another attribute potentially influencing the lipid profiles of lung tissue samples is the amount of residual blood in the samples.

For further study of lipid metabolic signatures in pathological events of human lung tissues, more samples have to be collected. This would enable to perfrom data analysis with a subset of the collected data, for instance only with males that were from 40 to 50 years old. This would reduce the complexity and enable to draw more confident conclusions. Sophisticated sampling techniques can be applied to study specific lipid metabolic signatures of fractions of the tissue samples. Primary AECII were successfully isolated from lung tissues before [165]. The analysis of specific lipid*ome* signatures from these cells would be of high interest in conjunction with the analytical procedures made before and could reveal whether the lipid metabolic processes

of AECII are affected with emphysema progression. Even other primary cells from lung tissues would provide valuable information in their lipid*omes* such as AMs or AECI although the last mentioned are not easy to isolate [166]. Moreover, AECII make an interesting and irreversible transformation to AECI in cell-culture [167]. Together with this de-differentiation, the ability of the cells to produce pulmonary surfactant gets lost. This would be an interesting process to monitor in their lipid profiles. However, when dealing with cell-cultures and primary cells, the protocols have to be carefully inspected. The presence of detergents and lipid containing media could cause unwanted disturbances for MS based lipid*omics* since detergents are known to suppress ionization and lipids in culture media could interfere with the cellular lipid*omes*.

3.3 PLS regression as tool for functional association in clinical tissue lipid*omics*

This study presented a PLS regression approach which made it possible to create a predictive model to evaluate multiple lipid*ome* influencing parameters. In the pilot study [146], all available lipid*omes* from either tumor or alveolar tissues were included in one model whose significance was tested by cv. Even with this limited number of cases, it was possible to associate the tissue morphology to specific lipid*ome* alterations. When later in the follow-up study more cases were included into the study, it enabled to improve the modeling strategy. Therefore, the complete cohort of collected lipid*omes* was 100 times randomly divided into training and validation subsets. The models computed with these 100 training subsets were cross-validated. Afterwards, the variability of the cv results provided information about the stability of the computed models. This methodology enabled identification of several lipid*ome* influencing parameters including gender, tumor entity and pulmonary emphysema as well as histopathological features.

Results from the regression models were cautiously interpreted due to the fact that the regressions were not sufficient to put them into predictive models as perspective for diagnostic applications as indicated by a validation utilizing the lipid*omes* from the independent validation subsets. The reasons for this limitations might be localized in the multidimensionality of the data. Most likely there are numerous unknown parameters further influencing lipid*omes*

of lung tissues such as smoking and diet which have not been considered in the models. Furthermore, the structure of the obtained histopathological characterizations most likely limited the predictive power of the computed regression models. Histopathological characterizations as shown here were manual categorizations of the tissue phenotype and not a continuous value that describes a histological feature. In particular, the scores for inflammation, AM and fibrosis were categorized into the stages 1, 2 and 3 which only provided a rough grouping of tissues with low, median and high degrees of these parameters. In addition, all these histopathological evaluations were partly subjective which was intercepted by the fact that the characterizations were always carried out by the same person with a specialist training as pathologist.

For future perspectives, it will be necessary to have completed data from the clinical archives. The histopathological scoring could be performed by multiple pathologists to evaluate the stability of the histopathological interpretations. The tissue morphology can potentially be quantitatively described by a continuous variable such as an absolute number of AM in a certain tissue volume and in more detail. Other analysis techniques might be used for the analysis of the tissue morphology such as automated image analysis which could reveal valuable information about the tissue phenotype that is not detected by manual interpretation. Probably, more sophisticated modeling approaches such as machine learning [168,169] should be applied in conjunction with increased case numbers. In future, this approach might be of high potential for development of novel diagnostic approaches for cancer differentiation, development of emphysema and COPD.

3.4 Future directions

There is a number of possible parameters in the methodological workflow that would improve data quality for high throughput analysis as well as data complexity. Further standardization in the reporting of lipid*omics* data is urgently required for rapid exchange of knowledge in the community. Therefore, possible improvements in data processing, data generation and data reporting are discussed in the following paragraphs.

Lack of bioinformatics tools for processing of lipid*omics* data is a current bottleneck of high throughput lipid*omics* studies. Software tools such as LipidXplorer [170] are suitable to extract signal intensities relevant for specific lipid species from mass spectrometric datasets.

Automated lipid assignments and proof by validation of specific fragmentation patterns is implemented. For further data processing, which includes data filtering, subtraction of background signals and quantification by referencing to ISDs, there is a lack of tools to automate these steps. Therefore, the processing of lipid*omics* data often requires time-consuming manual work. In this dissertation, the scripting language *R* was used to implement automated data processing and preparation procedures for multivariate statistical analyses. For these purposes, *R* provides a powerful toolbox for data processing and scripting allows to document each individual step. This enables high throughput processing of lipid*omics* data.

However, at the actual stage the scripts are not error tolerant and require a strict input data format. Furthermore, basic knowledge of the language *R* is required to run and debug the scripts. In future, these scripts can be used as base for development of an *'easy-to-use'* software toolkit for lipid*omics* data processing which might be made available to the scientific community.

Isomeric lipid species are a hidden treasure in lipid*omics* data. There are different levels, how deep lipid structures can be analyzed by MS. Lipid annotations generated in this study were analyzed to the level of sum compositions. Lipids were annotated with their sum of all aliphatic chains, such as PC 34:1 meant that both FAs together had 34 carbon atoms and 1 db. However, the assignment of PC 34:1 combines a large number isomeric species such as PC 16:0/18:1, PC 16:1/18:0 and PC 17:0/17:1, *etc*. In addition, there might be *sn* positional isomers as well as isomers with different db positions. There are methods available for mass spectrometric analysis of all of these isomers which might be routinely applied in lipid*omics* research in near future.

Today, there are still limitations in regard to the assignments the individual FA composition. Identification of individual FAs of a lipid molecule works sufficiently for GLs and GPLs with a maximum of two FAs per molecule, but fails for lipids containing more such as TAGs (3 FAs) and CLs (4 FAs). The db positions and *sn*-positional isomers can be determined by methods online coupled to MS such ozone-induced-dissociation (OzID) [171] and the Paterno-Büchi reaction (PB) [172]. At the actual stage of development, these methods are limited to be utilized on specific ion species and the most abundant signals due to the low sensitivity. However, the methodologies will improve in future and provide valuable information about lipid structures. Another emerging technique is ion mobility (IM) coupled to MS [173]. By determination of

molecular collisional cross-sections (CCS), isobaric species are separated by an additional dimension which further demands development of new data processing procedures. Interesting is that ratios of CCS and *m/z* differ for molecule classes, which means that this ratio is specific for lipids, peptides and carbohydrates. This enables acquisition of cleaner mass spectrometric data exclusively containing signals from lipids which increases analytical sensitivity and prevents false identifications. IM-MS is expected to be more sensitive than OzID and PB techniques since its sensitivity is not limited to a chemical reaction yields but it can be limited to specific ion species such as sodiated forms.

Standardization of data reporting allows sharing of lipid*omes* in public databases. Lipid*omics* is a relatively young discipline and standardization of experimental setups, data analysis strategies and data reporting is just at the beginning [174]. There are attempts for further standardization within the community [151]. However, integration of different platforms is not easy due to the individual demands regarding lipid nomenclature and data structure. The new lipid*omics* informatics for life science (LIFS) consortium aims to harmonize all steps of lipid*ome* analysis starting from the raw data processing up to storeage of their results in a public available database [175]. A comprehensive database combining lipid*omes* from various tissues, bio fluids and organisms will be a valuable resource for researchers. The lipid*omes* generated from this study will be made public available as general resource of the human lung lipid*ome*.

3.5 Conclusion

This work describes the characteristics of the lipid*omes* of human lung tissues. It is the first approach that provides a comprehensive lipid*ome* snapshot including the major GLs, GPLs, SLs and cholesterols and associates the lipid profiles to the histopathologic phenotype of the tissue samples and to the clinical parameters of the patients. The lipid*ome* compositions reflected the cellular reprogramming and differentiation in carcinogenesis enabling the discovery of new diagnostic markers and novel therapeutic approaches. The utilized modeling approach based on PLS regression revealed a complex pattern of lipid*ome* influencing parameters that shape the lipid profiles of tissue samples. This study showed that the histopathological phenotype has an immense effect on lipid metabolic processes and should always be considered when dealing with tissue samples. Additionally, patient related parameters such as age and gender induced specific alterations in the analyzed lipid profiles. The fact that these parameters are mirrored in the tissue lipid*omes* of patients could be of high interest for the personalized medicine concept. The impact of the patients' lifestyle parameters such as physical activity, nutrition, smoking or exposure to environmental toxins on lipid metabolic processes in the human lung is unknown but will reveal new insights into metabolic processes in pathological events.

4 Materials and Methods

4.1 Chemicals and lipid standards

Methanol, methanol containing 0.1 % (w/v) NH_4OAc, water, and 2-propanol were purchased from Fluka (Buchs, Switzerland) in LC-MS grade. Chloroform, MTBE, acetyl chloride (AcOCl), NH_4Cl, glyceryl triheptadecanoate (TAG 51:0), and human Hb were purchased in the highest available purity from Sigma-Aldrich (Munich, Germany). All other lipid standard substances were purchased from Avanti Polar Lipids (Alabaster, AL, USA) (Tabs. 4.1 and 4.2).

4.1.1 Preparation of internal standard solutions

Different preparations of ISD mix solutions were used for *sub-cohort 1* (pilot study [146]) and *sub-cohort 2*. SM d18:1/17:0, LPC 17:0, TAG 17:0/17:0/17:0, PE *OO*- 4Me 16:0/4Me 16:0, PC *OO*- 4Me 16:0/4Me 16:0, and Cholesterol d7 were present in both preparations (Fig. 4.1, Tab. 4.1). Therefore, these substances were used as references for lipid quantification to yield in comparable results between both subsets. The standard mix added to *sub-cohort 2* additionally included: PA 12:0/12:0, PG 12:0/12:0, PS 12:0/12:0, and CE 17:0 (Fig. 4.1, Tab. 4.2).

Abbreviations, as introduced in the tables 4.1 and 4.2, were used to name the standard substances.

Table 4.1: **Internal standards for lipid quantification and concentrations for samples from *sub-cohort 1*.**

Abbreviation	Substance	Supplier/ID	Conc. in ISD mix[a]	Conc. in sample[a]
SM-ISD	SM d18:1/17:0	Avanti/ 860585	465	0.93
LPC-ISD	LPC 17:0	Avanti/ 855676	654	1.31
TAG-ISD	TAG 17:0/17:0/17:0	Sigma/ T2151	392	0.78
PE-ISD	PE OO- 4Me 16:0/4Me 16:0	Avanti/ 999985	429	0.86
PC-ISD	PC OO- 4Me 16:0/4Me 16:0	Avanti/ 999984	407	0.81
Cholesterol-D7	Cholesterol-D7	Avanti/ 700041	847	1.69

[a] Concentration of the standard substance in pmol/µL.

Table 4.2: **Internal standard preparation used for *sub-cohort 2*.**

Abbreviation	Substance	Supplier/ID	Conc. in ISD mix[a]	Conc. in sample[a]
SM-ISD	SM d18:1/17:0	Avanti/ 860585	209	1.05
LPC-ISD	LPC 17:0	Avanti/ 855676	294	1.47
TAG-ISD	TAG 17:0/17:0/17:0	Sigma/ T2151	176	0.88
PE-ISD	PE OO- 4Me 16:0/4Me 16:0	Avanti/ 999985	193	0.97
PC-ISD	PC OO- 4Me 16:0/4Me 16:0	Avanti/ 999984	183	0.92
Cholesterol D7	Cholesterol-D7	Avanti/ 700041	1270	6.35
PA-ISD	PA 12:0/12:0	Avanti/ 840635	269	1.34
PG-ISD	PG 12:0/12:0	Avanti/ 840435	237	1.19
PS-ISD	PS 12:0/12:0	Avanti/ 840038	232	1.16
CE-ISD	Cholesteryl ester 17:0	Avanti/ 110864	234	1.17

[a] Concentration of the standard substance in pmol/µL.

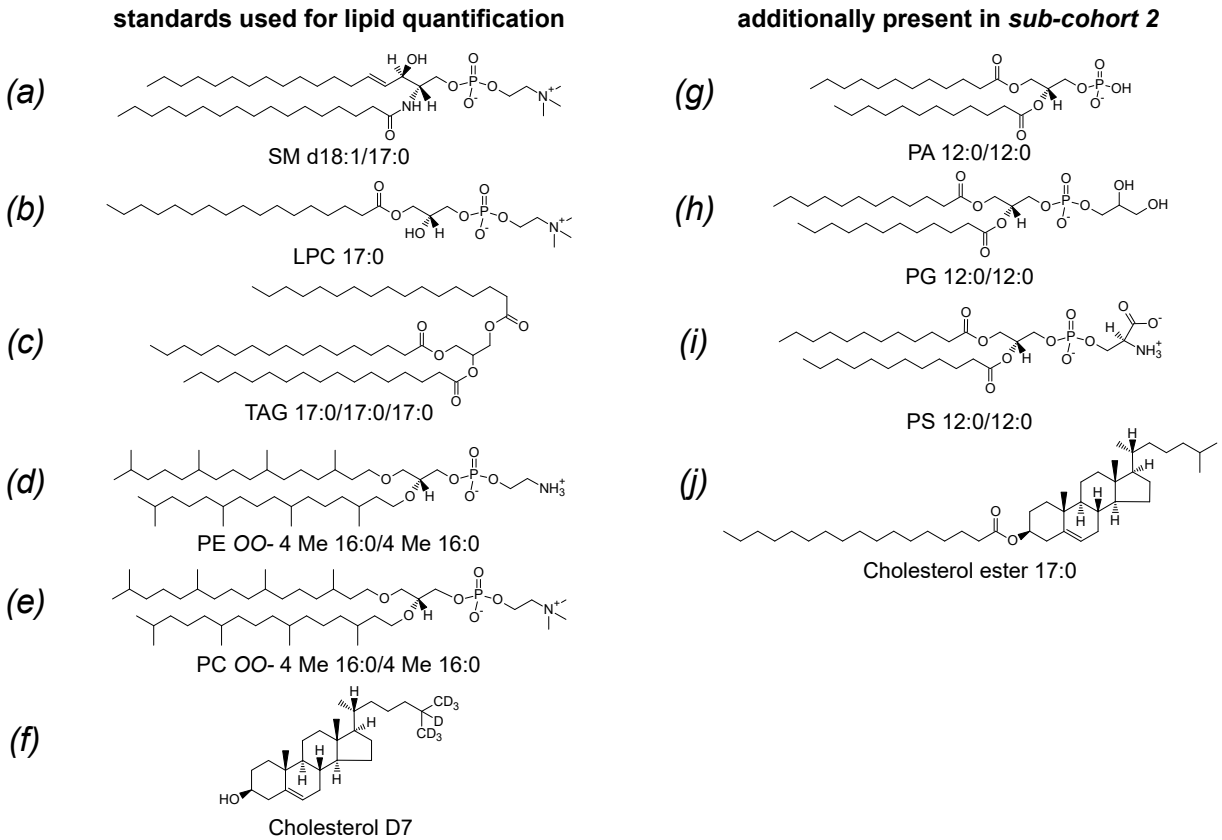


Figure 4.1: **Chemical structures of used internal standard substances.** Substances drawn on the left hand side were used as reference for all lipid quantifications. The lipid species drawn on the right hand side were also present in samples from *sub-cohort 2* but not further used for quantification.

Standard mix for *sub-cohort 1*

SM-ISD, LPC-ISD, TAG-ISD, PE-ISD, PC-ISD and cholesterol-D7 were purchased as powder (Tab. 4.1). Stock solutions with concentrations of 2 µg/µL were prepared in chloroform/methanol/water (60:30:4.5; v/v/v; storage solution) from each substance. Five hundred microliters of each of these six stock solutions were combined to yield in 3 mL of the readily prepared ISD mix solution. The solution was stored at −20 °C.

Standard mix for *sub-cohort 2*

Lipid standards were purchased as powder (SM-ISD, LPC-ISD, TAG-ISD, PE-ISD, PC-ISD, cholesterol-D7, PG-ISD, PS-ISD, and PA-ISD (Tab. 4.2)) and solved in the storage solution to prepare a 2 µg/µL stock solution for each. CE-ISD was supplied as solution with a concentration of 1 µg/µL solved in methylene chloride. Four hundred and fifty microliters of each, SM-ISD, PC-ISD, PE-ISD, LPC-ISD, PG-ISD, PS-ISD, PA-ISD and TAG-ISD, were combined with 900 µL CE-ISD and 1500 µL cholesterol-D7 to yield in 6000 µL of the ISD mix solution. Eleven aliquots, each with 500 µL of the solution, were prepared and stored at −80 °C.

4.2 Collection of human lung tissue samples

Human lung tissue samples were collected from March 2013 to January 2015. Samples originated from lung cancer patients who underwent lung cancer surgery at the *LungenClinic* (Großhansdorf, Germany). If available, one representative sample of the tumor and one sample of the peripheral tumor-free lung parenchyma, from here on termed as alveolar tissue, was taken from each patient in the size of 0.5 g to 1 g. Each lung tissue sample was divided into two parts to enable individual sample processing procedures for histopathology and shotgun lipid*omics* (Fig. 4.2, a). For shotgun lipid*omics*, the sample was shock frozen in liquid nitrogen and stored at −80 °C until use (Fig. 4.2, c). The other part was used for histopathological characterizations (Fig. 4.2, b). This work was conducted using material from lung cancer resections and performed anonymously. The use of patient tissues and all experimental procedures were approved by the local ethics committee of the University of Lübeck (AZ 12-220). All procedures were carried out in accordance with respective guidelines and regulations.

Table 4.3: **Clinical data of the included patients.**

	all collected tissue samples	***sub-cohort 1***	***sub-cohort 2***
No. of patients	92	26	66
No. of tissues	174 (86 A, 88 T)[a,b]	43 (21 A, 22 T)	131 (65 A, 66 T)
Age	40 - 84 (mean 63)	44 - 71 (mean: 58)	40 - 84 (mean: 66)
Female	32 (34 %)	6 (23 %)	26 (39 %)
ADC	42 (45 %)	11 (42 %)	31 (47 %)
SCC	31 (34 %)	12 (46 %)	19 (29 %)
Carcinoid	6 (7 %)	1 (4 %)	5 (8 %)
PY	10 - 125 (mean: 45, NA: 25)[c]	10 - 125 (mean: 47, NA: 2)	10 - 120 (mean: 43, NA: 23)
Never-smokers	4 (4 %)	0	4 (6 %)
BMI	16 - 40 (mean: 26, NA: 10)	18 - 40 (mean: 27, NA: 2)	16 - 40 (mean: 26, NA: 8)

[a] A: alveolar tissue
[b] T: tumor tissue
[c] NA: no data available

There were no initial selection criteria for the recruitment of patients. Two sub-cohorts were collected and combined for this work (Tab. 4.3). The first cohort was published as pilot study in Eggers *et al.* [146] comprising 43 lung tissues from 26 patients and termed *sub-cohort 1*. The second cohort included 132 tissue samples from 66 patients and was termed as *sub-cohort 2*. Combination of both cohorts was summarized as *all collected tissue samples* (Tab. 4.3).

In total, 174 tissue samples from 92 patients were collected. Paired samples from tumor-containing and tumor-free alveolar tissue were not available from each patient. Consecutively, 86 alveolar tissues and 88 tumor tissues were included (Tab. 4.3). From 92 patients, 32 (34 %) were female. Forty two (45 %) patients were diagnosed with ADCs and 31 (34 %) were suffering from SCCs. Moreover, 6 (7%) had carcinoid tumors. The remaining 10 patients were suffering from other classes of lung cancer such as LCCs, neuroendocrine tumors (NETs), or SCAs which represented minor groups. The ages of patients ranged from 40 to 84 years (mean 63 years). In addition, smoking behavior was recorded by the PY value [176]. PY is a statistical measure that describes the smoking rate of a patient and is calculated by multiplication of the number of cigarette packs (with 20 cigarettes) a patient smoked per day with the duration (in years) a patient smoked with this intensity. Most patients were smokers. For 63 patients, out of 92, PY values were recorded ranging from 10 to 125. A minor group of 4 patients comprised never-smokers. For 25 patients the smoking behavior was unknown. Furthermore, the BMI was recorded and ranged from 16 to 40 (mean 26).

4.3 Shotgun lipid*omics*

4.3.1 Tissue homogenization

Lung tissue samples were thawed on ice, transferred into sample tubes (Safe-Lock-Tubes, Eppendorf, Hamburg, Germany) and the tissue wet-weight was determined. Tissue samples (200-500 mg) were homogenized in 50 mM KCl-buffer using an Ultra-Turrax-Tube-Drive (IKA, Staufen, Germany). Therefore, the tissue samples were transferred into homogenization tubes for Turrax-Tube-Drive. For tissue amounts larger than 200 mg, DT-Tubes with rotor-stator elements were utilized. Tissue amounts lower than 200 mg were homogenized in BMT-Tubes that contained 25 stainless-steel balls. The ratio between tissue wet-weight and buffer volume was kept constant at 1:20 (w/v). This resulted in a concentration of 48 mg tissue per milliliter homogenate. A solution of butylhydroxy toluene (BHT) (1 mg/mL in methanol) was added to each sample from *sub-cohort 2* as anti-oxidant. To yield a constant BHT concentration per tissue wet-weight, 200 µL of the BHT solution were added per 1 g tissue. Tissues were homogenized for 2 min at 6000 rpm by inverse of the direction of rotation every 30 s. This step was repeated 2 to 3 times for DT-Tubes and 5 times for BMT tubes. To prevent the samples from heating up during homogenization, the tubes were stored in a fridge several minutes between repeated homogenization cycles. Four aliquots containing 300 µL homogenate were prepared, shock frozen in liquid nitrogen and stored at -80 °C. Three hundred microliters tissue homogenate contained 14.4 mg tissue (wet-weight).

4.3.2 Lipid extraction

Lipid extraction was based on the MTBE extraction protocol [105]. Three hundred microliters tissue homogenate were extracted according to the workflow scheme (Fig. 4.2, d). Tissue homogenates were thawed on ice and the ISD mix was spiked into the tissue homogenates. For *sub-cohort 1*, 4 µL ISD mix (Tab. 4.1) were used; for *sub-cohort 2*, 10 µL (Tab. 4.2). Afterwards, 2.25 mL of methanol were added to the samples. One blank extract was prepared for each extraction batch which contained 300 µL of the 50 mM KCl buffer, 2.25 mL of methanol and 3 µL of the BHT solution (1 mg/mL). After thorough mixing, 7.5 mL of MTBE were added and the mixtures were incubated for 30 min at room temperature while continuous shaking. Phase

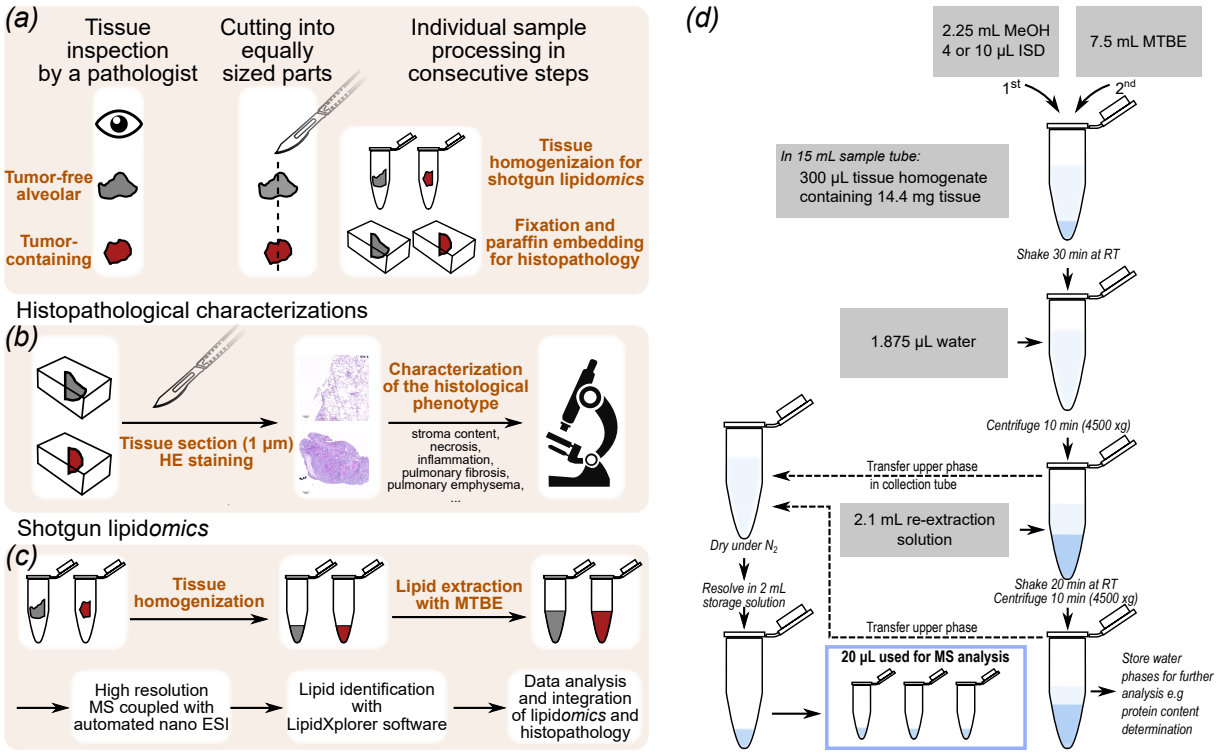


Figure 4.2: **Overview of the analytical workflow.** *(a)* After tissue inspection, tissue samples were cut into equally sized parts to enable histopathology and shotgun lipid*omics* with individual sample processing steps. *(b)* For histopathological characterizations, tissue sections were stained and characterized by light microscopy. *(c)* For shotgun lipid*omics* tissue samples were homogenized and lipids were extracted. *(d)* Lipid extraction procedure based on MTBE. Modified from Eggers and Schwudke (2017) [177].

separation was induced by addition of 1.875 mL of water. The samples were thoroughly mixed again, and centrifuged for 10 min at 4500 times the g-force ($\times g$). The upper phase (organic phase, mainly composed by MTBE and methanol) was taken off. Roughly, 1 cm supernatant of the organic phase remained above the water phases in the sample tubes. The upper phases were collected in glass tubes. The remaining water phases were re-extracted with 2.1 mL of the freshly prepared upper phase obtained by mixing MTBE, methanol and water 10:3:2.5 (v/v/v). After 20 min of incubation while continuous shaking, the mixtures were centrifuged for 10 min at 4500 $\times g$. The organic phases were collected and combined with the organic phases from the first extraction step. Again, a supernatant of 1 cm was left above the water phases in the sample tubes. The combined organic phases were dried under a nitrogen gas stream, dissolved in a mixture of chloroform, methanol and water (60/30/4.5; v/v/v), and stored at -20 °C.

4.3.3 Sample preparation for mass spectrometry

Lipid extracts, stored at −20 °C were removed from the freezer. Thereafter, the samples were carefully mixed to dissolve lipids that were potentially precipitated at low temperature. Twenty microliters of each sample and blank extracts, were diluted with 180 µL spray solution (dilution 1:10). In all cases, the spray solution was a mixture of isopropanol, methanol and chloroform (4/2/1, v/v/v). For positive ionization 3.7 mM NH_4OAc were added. Negative ionization was utilized with 0.05 mM NH_4Cl as additive. The samples were mixed and centrifuged for 5 min at 13,200 $\times g$. The samples (about 40 µL) were transferred to a 96-well plate (TwinTec, Eppendorf, Hamburg, Germany) and the plate was sealed with self-adhesive aluminum foil to prevent the solvent from evaporating.

4.3.4 *Top-down* shotgun lipid*omics* screens

Top-down shotgun lipid*omics* screens were performed on an Apex Qe (Bruker Daltonics, Bremen, Germany) FT-ICR mass spectrometer. The instrument was equipped with a TriVersa Nanomate (Advion Biosciences, Ithaca, USA) robotic and chip-based nano-ESI ion source. The Triversa Nanomate was utilized with D-Chips containing 400 nano-ESI nozzles with 4.1 µm inner diameter. The spray voltage was set to 1.1 kV and the back pressure to 1.1 psi. The Apex Qe was used with a nebulizer gas flow of 4.0 L/min, a dry gas flow of 1.3 L/min, and a dry gas temperature of 100 °C. The mass spectrometer parameters were set to a source accumulation of 0.01 s, collision cell accumulation of 0.7 s and a TOF time of 0.0014 s. Mass ranges were set from *m/z* 289.49 to 1200 in the positive ion mode and from *m/z* 289.49 to 1800 in the negative ion mode with a RP of approximately 100,000 at *m/z* 700. The MS^2 function was enabled with a collision energy of 2 V in positive ion mode and 0 V in negative ion mode. With this low collision energy no fragmentation is induced. The Q1 mass filter was set to a lower limit of *m/z* 400. Acquisitions were recorded over a period of 5 min.

Mass calibration was performed directly before starting the analysis sequence. The list for calibration comprised the ISDs such as CE 17:0 and SM 35:1 as well as compounds present in the sample matrix such as plasticizers, obtained from Schuhmann *et al.* [122] (Tab. 4.4). A consecutive list of polyethyleneglycol (PEG) polymers, often appearing as background ions, was calculated and included in the list for calibration (Tab. 4.4). Blank extracts were used for

Table 4.4: **Mass-to-charge-ratios of internal standards and chemical background compounds used for mass calibration.** List adopted from [122]. PEG: polyethylene glycol; xM: number of monomers/repeating units.

Calibrand	**Origin**	***m/z* (pos)** $[M+H]^+$	***m/z* (pos)** $[M+NH_4]^+$	***m/z* (neg)** $[M-H]^-$	***m/z* (neg)** $[M+Cl]^-$
PEG-6M	CB				317.1372
PEG-7M	CB	327.2014	344.2279		361.1635
Erucamide	CB	338.3417			
PEG-8M	CB	371.2276	388.2541		405.1897
PEG-9M	CB	415.2538	432.2804		449.2159
PEG-10M	CB	459.2800	476.3066		493.2422
PEG-11M	CB	503.3063	520.3328		537.2684
LPC 17:0	ISD	510.3554			544.3164
PA 24:0	ISD			535.3399	
Octadecyl (di-tert-butyl-hydroxyphenyl)propionate	CB			529.4626	565.4393
PEG-12M	CB	547.3325	564.3590		581.2946
PEG-13M	CB	591.3587	608.3852		625.3208
PG 24:0	ISD			609.3762	
PS 24:0	ISD			622.3715	
PEG-14M	CB	635.3849	652.4115		669.3470
Tris(ditert-butylphenyl)phoshite	CB	647.4588			
CE 17:0	ISD	656.6340			
Tris(ditert-butylphenyl)phoshate	CB	663.4537	680.4802		
Erucamide $[2M+H]^+$	CB	675.6762			
PEG-15M	CB	679.4111	696.4377		713.3733
SM 35:1	ISD	717.5905			751.5515
PEG-16M	CB	723.4374	740.4639		757.3995
PEG-17M	CB	767.4636	784.4901		801.4257
PE *OO*-	ISD	776.6892		774.6735	801.4257
PEG-18M	CB	811.4898	828.5163		845.4519
PC *OO*-	ISD	818.7361			852.6971
TAG 51:0	CB	866.8171			883.7516
PEG-19M	CB	855.5160	872.5426		889.4781
PEG-20M	CB	899.5422	916.5688		933.5044
PEG-21M	CB	943.5685	960.5950		977.5306
PEG-22M	CB	987.5947	1004.6212		1021.5568
PEG-23M	CB	1031.6209	1048.6474		1065.5830
PEG-24M	CB	1075.6471	1092.6737		1109.6092
PEG-25M	CB	1119.6733	1136.6999		1153.6355
PEG-26M	CB	1163.6996	1180.7261		1197.6617
Irganox 1010FF	CB		1194.8179	1175.7768	1211.753
PEG-27M	CB				1241.6879
PEG-28M	CB				1285.7141
PEG-29M	CB				1329.7403
PEG-30M	CB				1373.7666
PEG-31M	CB				1417.7928
PEG-32M	CB				1461.8190
PEG-33M	CB				1505.8452
PEG-34M	CB				1549.8714
PEG-35M	CB				1593.8977
PEG-36M	CB				1637.9239
PEG-37M	CB				1681.9501
PEG-38M	CB				1725.9763
PEG-39M	CB				1770.0025

mass calibration and for quality control regarding cleanness of the instrument and the connected parts.

4.3.5 Processing of mass spectrometric data

Raw spectra were automatically processed with Data Analysis 4.0 (Bruker Daltonics, Bremen, Germany). For each acquisition, high resolution survey scans were averaged from 1 min to 4.5 min acquisition time. Subsequently, the resulting averaged spectra were smoothed using Savitzky Golay method followed by automated peak picking. Finally, peak lists containing centroid *m/z* values, absolute intensities, peak FWHM, RP, relative intensities, and SNRs were exported as comma-separated-values (*.csv).

Next, the mass spectra quality was validated. Only acquisitions with base-peak intensities larger than 100,000 counts per single scan and stable spray conditions, monitored by the total ion current, were used. Additionally, only acquisitions were used where all ISDs were present. The data from the alveolar tissue of the patient with the ID 62 was removed since these criteria were not fulfilled.

4.3.6 Lipid identification

Exported *.csv files were imported into LipidXplorer software [170] version 1.2.7. LipidXplorer reads the first two columns from the peak lists: *m/z* and absolute intensity, other entries are ignored. To import and align multiple peak lists in LipidXplorer, instrument dependent data import settings were optimized (Tab. 4.5). The *m/z* range from 400 to 1000 was imported, which was the mass range in which signals from lipid molecules were expected. Calibration masses were utilized for internal mass re-calibration of each acquisition by masses from ISDs. LPC-ISD (*m/z* 510.3554) and PC-ISD (*m/z* 818.7361) were selected in the positive ion mode, and LPC-ISD (*m/z* 544.3175) as well as PE-ISD (*m/z* 774.6746) were selected for re-calibration in the negative ion mode. Mass tolerance was set to 10 ppm and peaks were filtered with an absolute intensity threshold of 500. Peaks were filtered by `min occupation` of 0.1. This value applied a filter and ensured that ions were present in at least 10 % of all acquisitions and excluded noise peaks. The resolution settings were estimated from a plot of the resolving power over the *m/z* (Fig. 4.3). The exponential decay of resolution with *m/z* in FT-ICR-MS

Table 4.5: **LipidXplorer import settings for import of data from an Apex Qe Fourier-transform ion cyclotron resonance mass spectrometer.**

Parameter	Value positive ion mode	Value negative ion mode
Selection window[a]	0.5 Da	0.5 Da
Time range[a]	0 - 1500 sec	0 - 1500 sec
Calibration masses MS	510.3554, 818.7361	609.3762,774.6735
m/z range MS	400 - 1000	400 - 1000
Resolution MS	140000	160000
Tolerance MS	10 ppm	10 ppm
Threshold MS	500	500
Resolution gradient MS	-150	-190
Min occupation MS	0.1	0.1
MS1 offset	0	0

[a] Arbitrary, value not needed

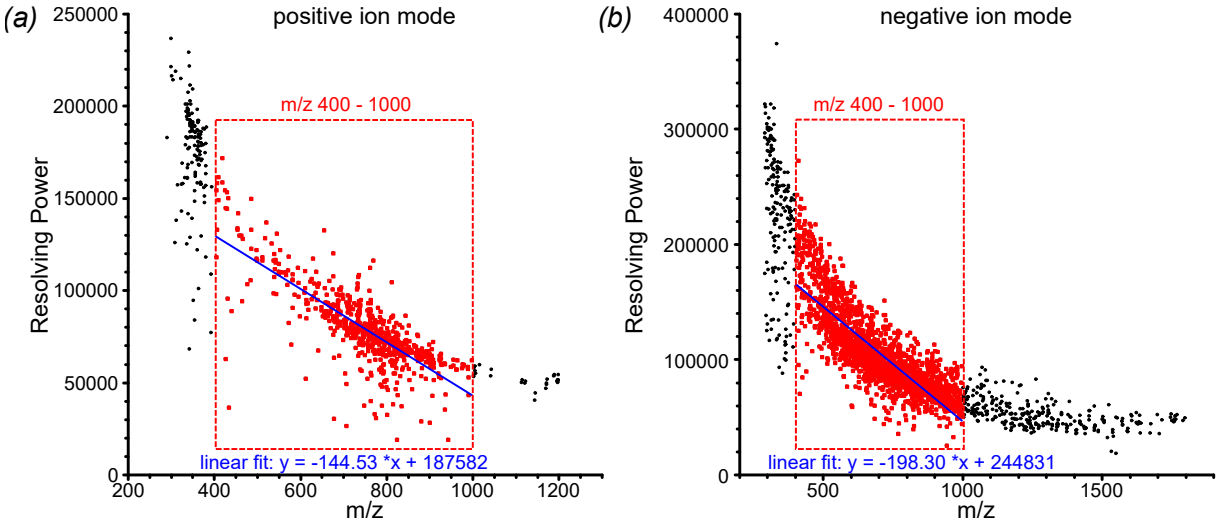


Figure 4.3: **Resolving power as function of the mass-to-charge ratio.** From example sample 25893-1. *(a)* Positive ion mode. *(b)* Negative ion mode. Highlighted in red is the *m/z* range from 400 to 1000 which was used for import in LipidXplorer. Blue shows a linear fit of this *m/z* region whose fit parameters are provided.

was approximated by a linear fit in the *m/z* region from 400 to 1000. The slope of the fit was set as `resolution gradient` in LipidXplorer and intercept at *m/z* 400 was used as `resolution` setting.

The resolution settings were optimized from a sample dataset comprising data from lung tissue extracts. Determined resolution gradients and resolutions at *m/z* 400 were averaged (Tab. 4.6). Rounded values with 2 significant digits were used as import settings. This yielded in a resolution at *m/z* 400 of 140,000 in positive and 160,000 in negative ion mode. The resolution gradient was set to -150 in the positive and to -190 in the negative ion mode.

Lipids were identified using customized scripts using the molecular fragmentation query lan-

Table 4.6: **Optimization of LipidXplorer import settings based on a sample dataset from lung tissue extracts.** (alveolar and tumor tissues) Sample ID: 25893, 25894, 25896, 25897, 25898, 25899, 25901, 25902, 25904 and 25905. Left: Settings from positive ion mode. Right: negative ion mode.

positive ion mode				*negative ion mode*			
Sample	**Intercept**[a]	**resolution**[b]	**resolution gradient**[c]	**Sample**	**Intercept**[a]	**resolution**[b]	**resolution gradient**[c]
25893-1	187581	129770	-144.53	25893-1	244831	165511	-198.30
25893-2	183950	127799	-140.38	25893-2	241456	165652	-189.51
25894-1	170361	120820	-123.85	25893-3	241845	165269	-191.44
25894-2	183695	126338	-143.39	25896-1	236617	165509	-177.77
25896-1	191803	131825	-149.95	25896-2	241775	167319	-186.14
25896-2	187237	128731	-146.27	25896-3	234008	163904	-175.26
25897-1	227037	159308	-169.32	25898-1	244659	164819	-199.6
25897-2	218539	152715	-164.56	25898-2	236688	164024	-181.66
25898-1	172758	122102	-126.64	25898-3	236321	163353	-182.42
25898-2	184869	126845	-145.06	25901-2	235405	162921	-181.21
25899-1	168803	118667	-125.34	25901-3	239981	164433	-188.87
25899-2	180674	125514	-137.90	25904-1	230541	160217	-175.81
25901-1	185360	127780	-143.95	25904-3	245604	167860	-194.36
25901-2	174777	122005	-131.93				
25902-1	343736	225180	-296.39				
25902-2	187611	128619	-147.48				
25904-1	231646	158494	-182.88				
25904-2	178307	123911	-135.99				
25905-1	200358	135130	-163.07				
25905-2	176151	123223	-132.32				
Mean	**196763**	**135739**	**-152.56**	**Mean**	**239210**	**164676**	**-186**
SD	*37864*	*23554*	*36.20*	SD	*4467*	*1874*	*8*

[a] at $m/z = 0$
[b] at $m/z = 400$
[c] m/z 400 to 1000

guage (MFQL) [170]. Utilizing these scripts, lipid species were assigned with a mass error less than 2.5 ppm and activated type-II isotopic correction. In the positive ion mode CE, DAG, LPC, LPE, MAG, PC, PC *O*-, PE, PE *O*-, PS, SM, TAG were screened. In the negative ion mode, Cer, CL, DAG, *lyso*-cardiolipin (LCL), LPA, LPC, LPE, *lyso*-phosphatidylglycerol (LPG), *lyso*-phosphatidylinositol (LPI), *lyso*-phosphatidylserine (LPS), PA, PC, PC *O*-, PE, PE *O*-, PG, PI, and PS were screened. Lipid identification was based on accurately determined *m/z* to calculate chemical sum compositions (section 1.3.5, page 34). Rules for lipid identification based on the chemical sum composition are given in table 4.7. It was searched for lipids with aliphatic chains with 12 to 24 carbon atoms and 0 to 6 dbs (Tab. 4.7). Exception: CE was screened for FAs up to 32 carbon atoms and DAG for up to 25 carbon atoms. CLs were identified by their $[M\text{-}2H+1]^{2-}$ ion (first isotopic peak), as proposed by Han *et al.* [178] since these ions have no overlap with other lipids such as PA or PG.

Table 4.7: **Rules for lipid identification based on the chemical sum composition.** DBR: Double bonds and rings.

		No. atoms in chemical sum composition						
Class	**Ion**	**O**	**N**	**P**	**C**	**H**	**Cl**	**DBR**
CE	$[M+NH_4]^+$	2	1	0	39 - 59	30 - 200	0	4.5 - 10.5
DAG	$[M+NH_4]^+$	5	1	0	27 - 54	20 - 200	0	0.5 - 12.5
LPC	$[M+H]^+$	7	1	1	20 - 32	43 - 67	0	0.5 - 6.5
LPE	$[M+H]^+$	7	1	1	17 - 29	37 - 61	0	0.5 - 6.5
MAG	$[M+NH_4]^+$	4	1	0	15 - 27	10 - 100	0	-0.5 - 5.5
PC	$[M+H]^+$	8	1	1	32 - 56	67 - 115	0	1.5 - 11.5
PC *O-*	$[M+H]^+$	7	1	1	32 - 56	67 - 115	0	0.5 - 10.5
PE	$[M+H]^+$	8	1	1	29 - 53	61 - 109	0	1.5 - 13.5
PC *O-*	$[M+H]^+$	7	1	1	29 - 53	61 - 109	0	0.5 - 12.5
PS	$[M+H]^+$	10	1	1	30 - 54	20 - 200	0	3.5 - 13.5
SM	$[M+H]^+$	6 - 7	2	1	31 - 49	67 - 103	0	0.5 - 7.5
TAG	$[M+NH_4]^+$	6	1	0	39 - 75	50 - 180	0	1.5 - 19.5
Cer	$[M+Cl]^-$	3 - 4	1	0	28 - 46	50 - 94	1	0.5 - 7.5
CL	$[M-2H+1]^{2-}$	17	0	2	56 - 104; ^{13}C: 1	90 - 250	0	5 - 24
DAG	$[M+Cl]^-$	5	0	0	27 - 51	30 - 130	1	1.5 - 13.5
LCL	$[M-2H+1]^{2-}$	16	0	2	44 - 80; ^{13}C: 1	50 - 250	0	4 - 18
LPA	$[M-H]^-$	7	0	1	15 - 27	20 - 80	0	1.5 - 7.5
LPC	$[M+Cl]^-$	7	1	1	20 - 32	43 - 67	1	0.5 - 6.5
LPE	$[M-H]^-$	7	1	1	17 - 29	33 - 57	0	1.5 - 7.5
LPG	$[M-H]^-$	9	0	1	18 - 13	20 - 80	0	1.5 - 7.5
LPI	$[M-H]^-$	12	0	1	21 - 33	10 - 140	0	2.8 - 8.5
LPS	$[M-H]^-$	9	1	1	18 - 30	10 - 200	0	2.5 - 8.5
PA	$[M-H]^-$	8	0	1	24 - 51	30 - 130	0	2.5 - 14.5
PC	$[M+Cl]^-$	8	1	1	32 - 56	66 - 114	1	1.5 - 11.5
PC *O-*	$[M+Cl]^-$	7	1	1	32 - 56	66 - 114	1	0.5 - 10.5
PE	$[M-H]^-$	8	1	1	29 - 53	57 - 105	0	2.5 - 12.5
PE *O-*	$[M-H]^-$	7	1	1	29 - 53	57 - 105	0	1.5 - 11.5
PG	$[M-H]^-$	10	0	1	27 - 54	30 - 120	0	2.5 - 9.5
PI	$[M-H]^-$	13	0	1	33 - 57	30 - 140	0	3.5 - 13.5
PS	$[M-H]^-$	10	1	1	30 - 54	20 - 200	0	3.5 - 13.5

Table 4.8: **Referencing of internal standards.** Quantities of underlined lipids were reported in the dimension absolute molar quantities [pmol]. *Italic* marked classes were reported in the dimension of *normalized abundances*.

Class	**ISD**	**Concentration in *sub-cohort 1* [pmol/mg tissue]**	**Concentration in *sub-cohort 2* [pmol/mg tissue]**
PC/*Cer*	PC-ISD	113.2	128.3
PE/*PS*/*PG*/*PA*/*PI*/*CL*/*Lyso*	PE-ISD	119.3	135.3
LPC	LPC-ISD	181.7	206.0
TAG/*DAG*/*MAG*/*CE*	TAG-ISD	109.0	123.6
SM	SM-ISD	129.1	146.4

Lipid species were annotated according to the short-hand notation proposed by Liebisch *et al.* [179]. Different from the short-hand notation, the number of additional hydroxylations (OHs) within SLs (Cer, HexCer, SM) was additionally noted after a semicolon (;). For example a species Cer 34:1;0 would carry 34 carbons and one db in the sphingoid base and the FA residues. Assuming a C_{18} sphingosine base, the molecule carries one hydroxylation in the 1-position and one hydroxylation in the α-position to the amide whereas Cer 34:1;1 carries one additional OH in the molecule. The position is not specified with this nomenclature.

4.3.7 Lipid quantification and post-processing

Lipid species were quantified preferably based on signals from a ISDs substance from the same class. For lipid classes without a class-specific ISD present in the samples, signals from ISDs belonging to another lipid class were selected. Therefore, different strategies were utilized for the data from the *pilot study* (section 2.2, starting at page 48) and from the *follow-up study* (section 2.3, starting at page 59).

In the *pilot study*, the mass spectrometric signals of lipid species without a class-specific ISD were referenced to the sum of the signal intensities of LPC-ISD, SM-ISD, PC-ISD and TAG-ISD in the positive ion mode. In the negative ion mode, these signals were referenced to the sum-intensity of LPC-ISD, PC-ISD, and PE-ISD [146].

For lipid quantification as utilized in the *follow-up study*, lipid species without a class-specific ISD were referenced to a single reference compound, since this simplified automated data processing as it was scripted utilizing the language *R*. Therefore, a representative standard substance was chosen which undergoes similar adduct forming mechanisms during the ESI process

Table 4.9: **Lipid species that were removed from the dataset due to invalid lipid identifications.**

Lipid assignment	*m/z*	Polarity	Reason for deletion
PE 31:7	662.3812	negative	no isotopic distribution
Cer 42:4;0	678.5582	negative	noise
Cer 43:6;0	688.5447	negative	noise
Cer 43:5;1	706.5537	negative	noise
PE *O*- 37:5	736.5281	negative	no isotopic distribution
Cer 46:5;1	748.601	negative	noise
PE *O*- 39:7	760.5281	negative	noise
PG 37:4	783.5188	negative	noise
PA 47:10	795.497	negative	noise
PA 44:10	795.499	negative	artefact
PE *O*- 42:10	796.5306	negative	artefact
PA 44:5	805.5761	negative	isotope
PE *O*- 44:10	824.562	negative	no isotopic distribution
PC *O*- 37:0	824.631	negative	isotope
PA 46:6	831.5891	negative	artefact
PC *O*- 39:8	836.5376	negative	noise
PE *O*- 45:4	850.6676	negative	isotope
PC *O*- 39:0	852.6638	negative	isotope
PA 48:9	853.5757	negative	noise
PG 43:6	863.5795	negative	artefact
PE 45:4	864.6467	negative	artefact
PC *O*- 42:9	876.5696	negative	no isotopic distribution
PC *O*- 43:9	890.5836	negative	isotope
PI 39:3	901.5813	negative	noise
PS 47:0	944.7327	negative	peak profile

as it is summarized in table 4.8. Cers were referred to the PC-ISD since both form $[M+Cl]^-$ ions in the negative ion mode. PS, PG, PA, CL and their corresponding *lyso* species were grouped together and all referred to the PE-ISD since they ionize as $[M-H]^-$ ions. In the positive ion mode, DAGs, MAGs, and CEs were grouped together and referred to the TAG-ISD since all form NH_4^+ adduct ions. As consequence, the quantities of lipid classes with a class specific ISD substance were reported in molar scale, *e. g.* as pmol values. Quantities of those classes without respective class-specific ISDs were reported as *normalized abundances*.

The raw outputs from LipidXplorer were further revised to remove chemical noise. Furthermore, species with same exact masses were removed by deleting all lipids that were annotated by the `isobaric` tag in LipidXplorer [30]. The chemical background was removed from the datasets. Lipids with average quantities across all samples lower than 10-fold of the average quantity determined in blank extracts were removed from the datasets. Probably, these lipids were background noise signals and false-positive identified lipids. Results from positive and

negative ion modes were combined to the analyzed lipid*omes*. The lipid classes CE, DAG, MAG, SM, and TAG were preferably quantified in the positive ion mode. Cer, CL, PA, PC, PE, PG, PI, PI, and their *lyso* species were quantified in the negative ion mode. Lipid quantification, background subtraction, and merging positive and negative ion mode results was scripted using *R* in the follow-up study (The *R* Project [180]).

Afterwards, the lipid*omes* were manually revised for lipids with an uncommon sum-composition. In particular, there was a strong focus on lipids with short-chain PUFAs and lipids with long carbon chains, which were highly saturated. The signals of these lipids were manually validated by analysis of isotopic distributions and peak profiles (Fig. 4.4, Tab. 4.9). It was evaluated whether the respective peak was an isotopic signal from another signal where a mass difference of 1.003 amu was expected which is the excact mass difference of the ^{12}C and ^{13}C isotopes. For example, the signal at *m/z* 890.5833 was identified as an isotopic peak of the signal at *m/z* 889.5799 (Fig. 4.4, a). The raw LipidXplorer outputs assigned this signal as PC *O*- 43:9. Moreover, isotopic distributions were validated to fit to predicted patterns expected for lipid molecules. For the peak at *m/z* 662.3815 no isotopic distribution was present in the respective mass spectra (Fig. 4.4, b). Potentially, this peak was a spike signal which originated from electronic noise and not as assigned the lipid species PE 31:7. The peak profiles were validated in the next step. For example, the intensive peak at *m/z* 863.5649 showed two small satellite peaks from which one was identified as PG 43:6 (Fig. 4.4, c). These satellite peaks were most likely artifacts from the spectra processing by Fourier-transformation, so-called *sinc wiggles*, which appear with signal truncation within the recorded transient. All false positive lipid assignments were deleted from the datasets (Tab. 4.9). After the manual revision, technical replicates for each sample were averaged.

4.3.8 Optimization of the ammonium chloride concentration for negative ionization

Five spray solutions (chloroform/methanol/2-propanol; 1:2:4; v/v/v) were prepared containing different concentrations of the NH_4Cl salt additive of 0.5 mM, 0.25 mM, 0.1 mM, 0.05 mM and 0.01 mM. Four lipid extracts from alveolar tumor-free human lung tissue samples (Sample No. 26417, 26418, 26419, 26420, 26421) and were diluted 1:10 with each spray solution and

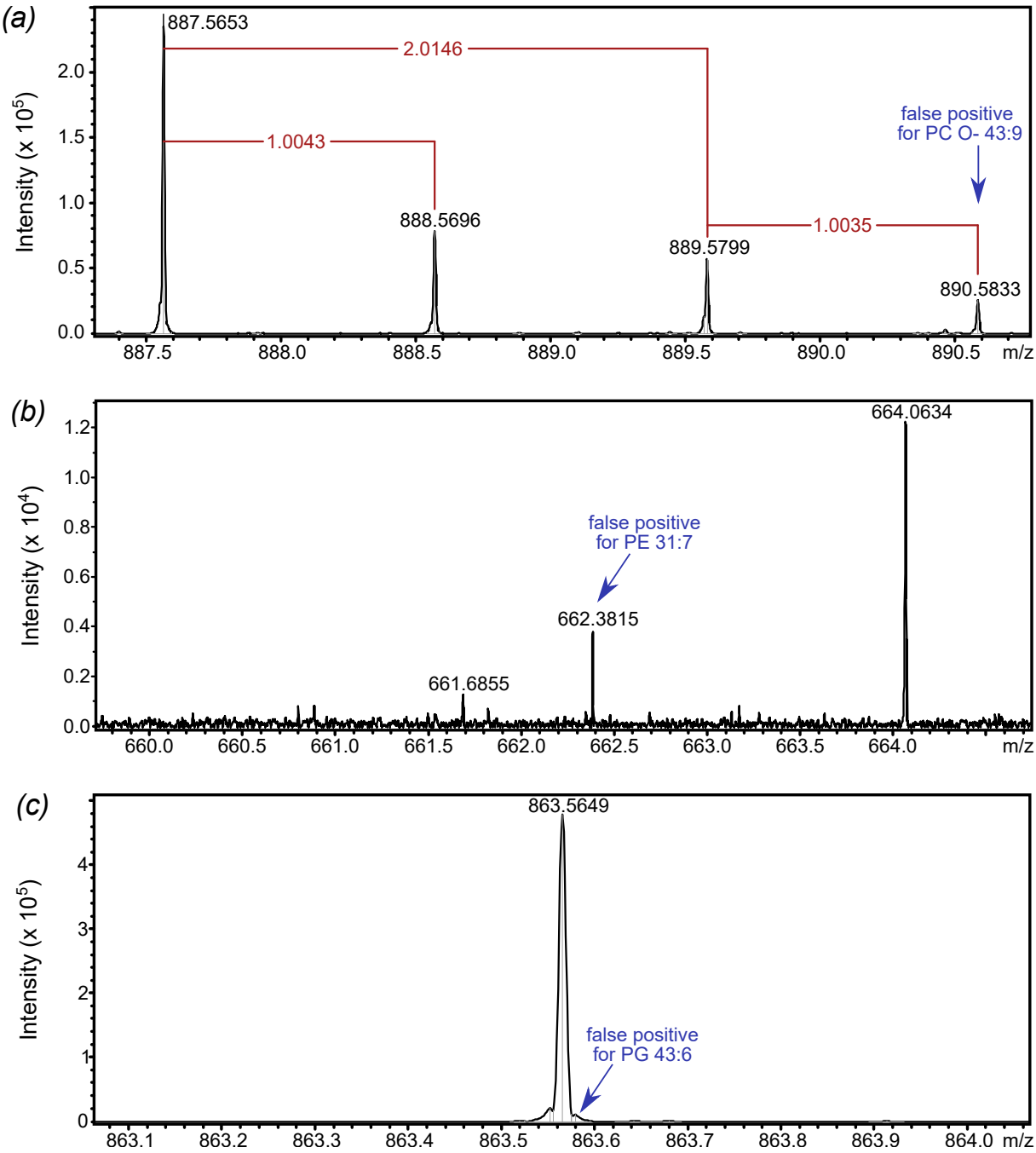


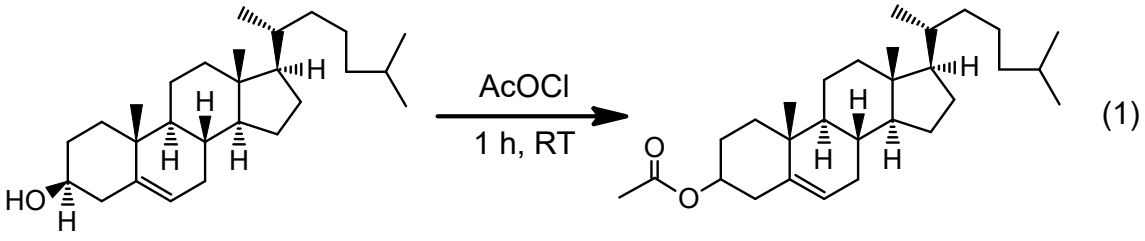
Figure 4.4: **Decision criteria for valid mass spectrometric signals.** *(a)* Isotopic distributions were evaluated. Is the peak an isotopic peak of another signal? *(b)* Does the peak have a proper isotopic distribution, or is it an artifact? *(c)* Peak profiles were validated. Is the peak an artifact from Fourier transformation?

analyzed by FT-ICR MS. ISDs were used (Tab. 4.1). Spectra were processed as described for shotgun lipid*omics*. SNRs instead of signal intensities were imported into LipidXplorer software. After lipid identification, the results were filtered: only lipid species, which were present in all measurements, except measurements of blank extracts, were used for further analysis. SNRs were summed up for each lipid species of a respective lipid class and the summed values were normalized to the average values obtained an average NH_4Cl concentration of 0.5 mM. Values processed in this manner were termed as PCSNR. Additionally, the SNRs of the ISD signals of PC-ISD, LPC-ISD and PE-ISD were normalized to their respective value at an NH_4Cl concentration of 0.5 mM.

4.3.9 Quantification of free cholesterol

Derivatization of free cholesterol with acetyl chloride

FC was quantified according to the protocol published by Liebisch *et al.* [181]. FC is an unpolar molecule that hardly ionizes during the electrospray process. The method by Liebisch *et al.* derivatizes FC with AcOCl to form cholesterol acetyl ester as shown in reaction scheme (1).



free cholesterol cholesterol acetyl ester (1)

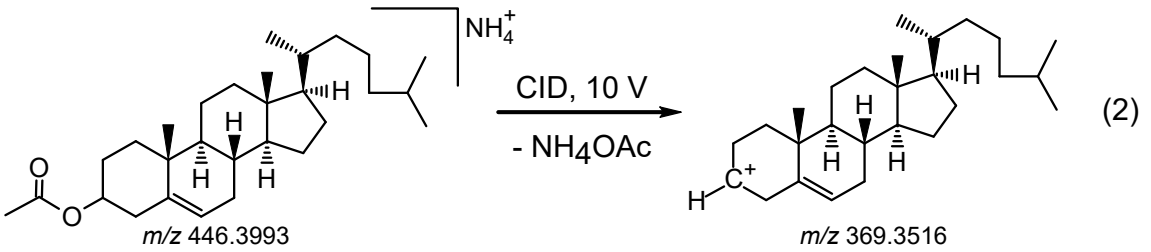
The lipid extracts (250 µL) were transferred into separate sample tubes (Safe-Lock Tube, 2 mL, Eppendorf, Hamburg, Germany) and concentrated until complete dryness in a vacuum concentrator (Savant SpeedVac, Thermo Fisher). AcOCl was diluted 1:5 (v/v) with chloroform in a pre-washed glass tube. One hundred microliters of this solution were added to each dried sample. The mixture was incubated for 60 min at room temperature in open sample tubes. In the last step, the remaining solvent was removed in the vacuum and the dried samples were stored at -20 °C.

Free cholesterol quantification by quadrupole-time of flight tandem mass spectrometry

Derivatized and dried lipid extracts were solved in 200 µL of a mixture of chloroform, methanol and isopropanol (1/2/4, v/v/v) containing 3.7 mM NH_4OAc. Samples were mixed thoroughly and centrifuged for 5 min at 13,500 $\times g$. The eluent was prepared by mixing 150 mL chloroform, 300 mL methanol containing 0.1 % (w/v) NH_4OAc and 600 mL isopropanol in a glass bottle. Afterwards, the mixture was degassed for 15 min in an ultrasonic bath.

Samples were analyzed on a Q-TOF hybrid mass spectrometer, QTOF Ultima (Waters, Milford, USA), utilizing automated flow-injection for sample infusion. A 1100 series high pressure liquid chromatography (HPLC) system (Agilent Technologies, Waldbronn, Germany) without connected chromatographic column was used for sample infusion. The HPLC system consisted of a degasser, a binary capillary pump and an autosampler equipped with a cooling unit. The capillary pump was set in micro flow mode. One mircroliter sample volume was injected into an isocratic flow of the eluent with a flow rate of 8 µL/min. The flow rate was kept constant at this flow until 2.5 min after injection. Afterwards, from 2.5 min to 4.5 min post injection the flow was increased to 20 µL/min to wash the system. From 4.5 min to 5 min post injection the flow rate was equilibrated back to 8 µL/min for the next run. The autosampler thermostat was set at 15 °C.

The HPLC system was directly coupled to the ESI source of the Q-TOF mass spectrometer with a poly-ether-ether-ketone (PEEK) capillary. Ions were detected in positive ion mode by successive MS^2 scans in parallel reaction monitoring mode with a precursor selection window of about 1 Da. Fragmentation was induced by CID with collision energy of 10 V. Fragmentation of cholesterol acetyl ester is illustrated in reaction scheme (2):



(2)

Cholesterol acetyl ester, which was generated in the derivatization reaction with AcOCl and selected as precursor ion at *m/z* 446.5 ($[M+NH_4]^+$). Cholesterol-D7 acetyl ester, generated from cholesterol-D7 during the derivatization, was used as ISD for quantification of FC and

monitored with a precursor mass of *m/z* 453.5 ($[M+NH_4]^+$). Full MS^2 spectra were recorded for each precursor with a scan time of 1 s and an inter-scan delay of 0.1 s and only centroid *m/z* values were recorded.

The resulting datasets were analyzed using the *QuanLynx* module from MassLynx 4.0 software (Waters, Milford, USA). Extracted ion chromatograms within a mass window of 0.5 Da were generated for the transitions *m/z* 446.5 to 369.5 (FC) and *m/z* 453.5 to 376.5 (cholesterol-D7). The flow-injection peak appeared between 0.5 and 1.5 min post injection and was automatically integrated utilizing the *QuanLynx* software. The integration was manually revised for each sample. The integration results were exported as tab-delimited text (*.txt) file which was afterwards imported into Microsoft Excel. FC was quantified by referencing its peak area to the peak area of cholesterol-D7, and multiplication with the known cholesterol-D7 concentration. Quantities of FC were reported as molar amount (pmol) per milligram tissue wet-weight.

4.4 Histopathological characterization of lung tissues

All tissue preparations were performed in the Clinical and Experimental Pathology of the Research Center Borstel. Tissue samples used for histopathological characterizations were directly fixed after sampling using the HOPE technique as described in detail by Olert *et al.* [90]. Fixed tissue samples were dehydrated with acetone and embedded into a paraffin block. For analysis, tissue sections of 1 µm were prepared and stained with HE. Afterwards, tissue slices were analyzed by light-microscopy. All histopathological characterizations and were performed by Dr. Julia Müller who is a medical specialist in pathology.

4.4.1 Histopathological scoring of tumor tissues

Tumor tissues were classified into categories according to the tumor type: ADC, SCC, LCC, carcinoid, SCA, and NET. With regard to the tissue morphology, the phenotype of tumor tissues was expressed in histopathological scores. The relative composition of the tissue samples was scored as tissue fraction. The tissue fraction (in percent) gives the relative proportions of metabolic active tumor cell, stroma compartments and necrotic areas. In sum, these parameters were defined to results always in 100 %. Furthermore, the infiltration by immune cells was

expressed in an inflammation score which categorizes tissues into inflammation stages 0, 1, 2, 3 from low to high.

4.4.2 Histopathological scoring of alveolar tissues

Due to their distinct morphology, different criteria were applied for histopathological scoring of alveolar lung tissues. The main phenotype of alveolar lung tissues is the pulmonary emphysema which is an irreversible and pathological enlargement of the alveolar structures. This overall enlargement of the alveoli is visible on HE stained tissue slices. Using the method proposed by Nagai *et al.* [148], the tissues were categorized in emphysema stages from 0 to 10, where 0 means a normal alveolar structure and 10 a complete loss of the alveoli. The inflammation status was determined analogue as performed in tumor tissues. Additionally, pulmonary fibrosis and infiltration by AMs was categorized in stages 0 to 3, from low to high.

4.5 Quantification of hemoglobin content in tissue homogenates

Aliquots of 300 µL lung tissue homogenate were thawed. Alveolar tissue homogenates were diluted with 300 µL KCl buffer (50 mM). Tumor tissue homogenates were used undiluted. Each sample was centrifuged for 1 minute at 4500 $\times g$ until a clear supernatant remained. Aliquots of 100 µL from these supernatants were transferred onto a 96-well plate for spactralphotometric analysis. Absorbance spectra were recorded from 320 nm to 700 nm with a step size of 2 nm on an Infinite 200 Pro plate reader (Tecan, Maennedorf, Switzerland). Hb standard substance (Sigma-Aldrich, Munich, Germany) was used for calibration and quantification. Calibration solutions were prepared with 1 mg/mL, 0.5 mg/mL, 0.25 mg/mL, 0.125 mg/mL and 0.0625 mg/mL Hb content in 50 mM KCl buffer. One hundred microliters of the calibration solutions were pipetted on the 96-Well plate. Each sample and standard was analyzed in duplicate.

Due to the coordination of iron-II ions in the hem-complex of Hb, absorbance spectra show an intense absorption band with a maximum at 406 nm (Fig. 4.5, a). Integration of this band from 370 to 450 nm enabled a linear range for quantification from 0.0625 mg/mL up to 1 mg/mL

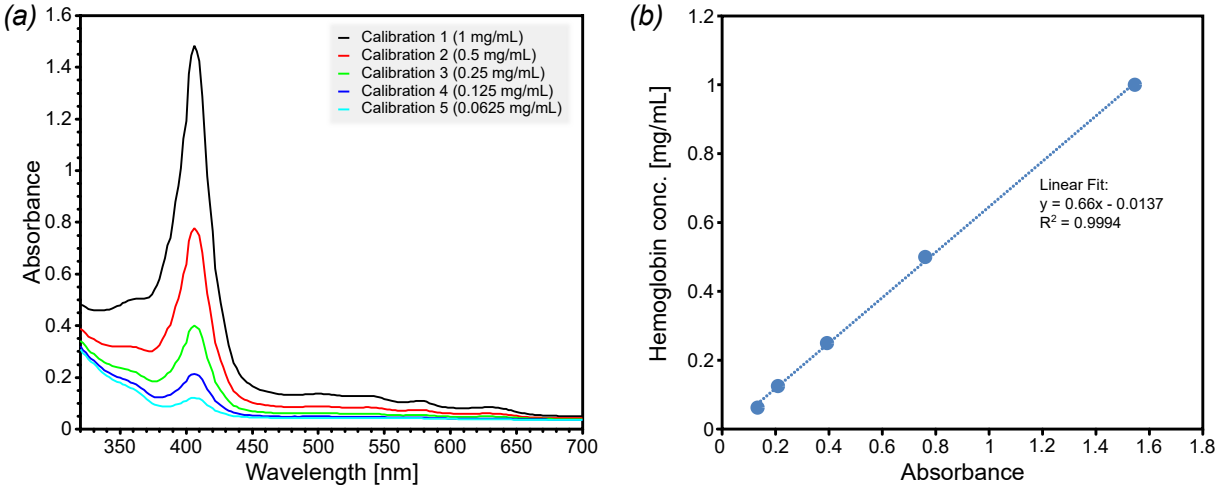


Figure 4.5: **Spectral-photometric quantification of hemoglobin in lung tissue homogenates.** *(a)* Absorbance spectrum of Hb standards from 320 to 700 nm. *(b)* Calibration curve for Hb from 0.0625 mg/mL to 1 mg/mL.

Hb concentration (Fig 4.5, b). Spectral integration was performed with the software *SciDaVis*.

4.6 Statistical data analysis

4.6.1 Data normalization

All lipid quantities were expressed as *normalized abundances* (section 4.3.7, page 111). Afterwards, each lipid abundance was normalized to the sum of all *normalized abundances* per tissue lipid*ome* (relative abundance). Therefore, the sum of all *relative abundance* off all lipid species in a lipid*ome* is always 100 %. In the pilot study, normalized abundances were calculated based on the complete set of identified lipid species. In the follow-up study, normalization was performed after application of data filters.

4.6.2 Lipid*ome* homology based on LUX scores

The calcultation of LUX scores was part of the pilot study [146]. Transformation of lipid structures into the chemical space model and lipid*ome* homology based on LUX scores was calculated by Dr. Chakravarthy Marella as described previously [145]. A list of FAs which were present in the utilized lung tissue extracts was determined by MS^2 experiments performed on a Q Exactive Plus (Thermo Scientific, Bremen, Germany) Orbitrap mass spectrometer in the

Table 4.10: **Fatty acids used to draw lipid structures for LUX score calculation.** db: carbon-carbon double bond. Table from [146].

0 db	1 db	2 db	3 db	4, 5 and 6 db
0:0	14:1:(9Z)	16:2:(9Z,12Z)	18:3:(9Z,12Z, 15Z)	20:4:(9Z,12Z, 15Z, 18Z)
14:0	16:1:(9Z)	17:2:(9Z,12Z)	20:3:(9Z,12Z, 15Z)	22:4:(9Z,12Z, 15Z, 18Z)
15:0	17:1:(9Z)	18:2:(9Z,12Z)	22:3:(9Z,12Z, 15Z)	24:4:(9Z,12Z, 15Z, 18Z)
16:0	18:1:(9Z)	20:2:(9Z,12Z)	24:3:(9Z,12Z, 15Z)	20:5:(6Z, 9Z,12Z, 15Z, 18Z)
17:0	19:1:(9Z)	22:2:(9Z,12Z)		22:5:(6Z, 9Z,12Z, 15Z, 18Z)
18:0	20:1:(9Z)	24:2:(9Z,12Z)		22:6:(6Z, 9Z,12Z, 15Z, 18Z, 21Z)
19:0	21:1:(9Z)			
20:0	22:1:(9Z)			
22:0	24:1:(9Z)			
24:0				

negative ion mode. A list of FAs is noted in table 4.10. A list of 311 identified lipid species, including FC, from the 43 tissue lipid*omes* of *sub-cohort 1* (Tab. 4.3) was used as input data. Chemical structures of 293 lipid molecules were generated out of the 311 identified lipids. A list of 18 lipids was not included into the analysis due to the fact that no combination of FAs from table 4.10 fitted to the respective sum composition. Lipid*ome* maps were calculated with *pca* based on structural similarities of the lipid molecules with the *R* function `princomp` from the `stats` library [180]. Lipid*ome* homology was visualized by calculation of a hierarchical clustered tree based on pairwise LUX scores between each comparison of tissue lipid*omes*. Hierarchical clustering was calculated with the *R* function `hclust` from the `stats` library using *Euclidean distance* as metric and *complete linkage* as clustering method. Error modeling was performed to test the robustness of the hierarchical clustered tree. Therefore, a threshold for quantitative values was defined to exclude the lowest concentrated lipid species from the analysis. This respective value should simulate the detection threshold (*dt*) of the analytical system. To validate the robustness of the hierarchical clustering results, all quantitative values were varied randomly with a pre-defined standard deviation (SD) for 100 iterations. This procedure was performed for $dt = 0.005$, SD $= 0.002$ and $dt = 0.003$, SD $= 0.001$.

The matrix of LUX scores was further analyzed by *pca*, calculated using the `PCA` function of the *R* library `FactoMineR` [182]. Afterwards, the coordinate of pc_1 was used as discriminator for alveolar and tumor tissue lipid*omes*. A ROC [183] curve was calculated based on pc_1 and the tissue classification (alveolar or tumors) as input data. The ROC curve was calculated and visualized by Graph Pad Prism 6 (GraphPad Software Inc.).

4.6.3 Hierarchical clustering and heat map

Pilot study

Hierarchical clustering was calculated with the lipid*ome* dataset from *sub-cohort 1* (pilot study) which included 311 lipid species within 43 tissue lipid*omes* [146]. Lipid*ome* data in dimension of normalized abundances were used as input data and further processed using Gene Cluster 3.0 [184]. The parameter `%Present` was set to 90 % as data filter which resulted in a dataset comprising 141 lipid species that were present in at least 90 % of all 43 samples. The data were adjusted by enabled `Log tranform data`, `Center genes` and `Center arrays` check boxes which achieved data that were log2 transformed and centered. Afterwards, hierarchical clustering was calculated with `Euclidean distance` as similarity metric and `Complete linkage` as clustering method by the software. The clustering results were visualized using Java TreeView 1.1.6r4 [185].

Follow-up study

For the follow-up study, the combined data from *sub-cohort 1* and from *sub-cohort 2* were used which finally included data from 174 lipid*omes*. The lipid*omes* were filtered in the first step for occupation of lipid species across all samples. One hundred and forty five lipid species were included, present with a quantitative value in at least 80 % of the 174 samples. After normalization as relative abundances, *fold-changes* were calculated per lipid species, compared to its average relative abundance, and the data were log2 transformed afterwards. A hierarchical clustering with heat map representation of individual lipid abundances was calculated with the *R* function `heatmap.2` from the `ggplot2` package [186] utilizing *Euclidean distance* as similarity metric and *average linkage* as clustering method.

4.6.4 Principal component analysis

The same dataset as used for hierarchical clustering with combined data from *sub-cohort 1* and *sub-cohort 2* was analyzed. The data were filtered with an occupation cut-off of 80 % resulting in a dataset with 145 lipids from 174 tissue lipid*omes*. Missing values were imputed using the `imputePCA` function from the *R* library `missMDA` [144]. This function uses an iterative *pca*

algorithm to impute missing values [144]. Quality control of the imputed data was performed by multiple imputations with the `MIPCA` function from the `missMDA` package for 50 iterations (data not shown). Multiple iterative imputations demonstrated robustness. Based on the imputed dataset, *pca* was calculated using the `PCA` function from the *R* package `FactoMineR` [182].

4.6.5 Statistical tests comparing tissue classifications

Group wise comparison of alveolar and tumor tissues. The same input data was used for grouped testing between alveolar and tumor tissues. Multiple unpaired U-tests (*R* function `wilcox.test`) were calculated to compare alveolar and tumor tissues. Afterwards, *p*-values were adjusted by using Benjamini-Hochberg FDR using the `p.adjust` function with activated `BH` option from the `stats` *R* library. The results were visualized in a volcano plot where the adjusted *p*-values were plotted over the log2 fold-changes of group wise mean values per lipid species. The input data were filtered by adjusted *p*-values and log2 fold-changes. Lipids with log2 fold-changes below -1 or larger than +1, and with adjusted *p*-values lower than 0.05 were considered for further analysis. A hierarchical clustering exclusively based on these lipids was calculated (section 4.6.3, page 121). Abundances of the selected lipids were summed-up to a score for tissue discrimination. To calculate the respective score, the selected quantitative values were transformed as log2 fold-changes compared to the overall mean per lipid specie. Next, each value was multiplied with the corresponding log2 fold-change of means per lipid species (single value per species). In the last step, the sum over all lipids per tissue sample was calculated to yield in one discriminative value per sample. The discriminative score was evaluated for its diagnostic properties in a ROC curve. All ROC curves were calculated and visualized by using the `plot.roc` function from the *R* package `pROC` [187].

Group wise comparison of lipid*omes* from tumor subsets. In the same manner, subsets of the tumor lipid*omes* were analyzed. Lipids present in at least 97 % of all samples of the respective comparison were included. The lipid*omes* that were identified as outliers in previous analyses (ID24_T and ID125_T) were excluded. ADC lipid*omes* were compared with SCC lipid*omes*. For comparison of carcinoid tumor lipid*omes* with NSCLC tissue lipid*omes*, SCC, ADC, and LCC lipid*omes* were grouped together as NSCLC tissues. The cut-off for the log2 fold-changes was set to +0.5 and to -0.5, respectively.

Group wise comparison of lipid*omes* from alveolar subsets. The workflow as described in the paragraph before was further applied to compare lipid*omes* of the tumor-free alveolar tissues from ADC and SCC patients. The tissue ID101_A was excluded from this analysis. Consequently, alveolar lipid*omes* from 39 ADC patients and from 28 SCC patients were included into this analysis. From this subset, all lipid species were included into the analysis that were present in at least 80 % of the samples. The cut-off for lipid selection by the respective log2 fold-change was set to +0.1 and -0.1.

4.6.6 Partial least-squares regression analysis

Pilot study

In the pilot study, PLS regressions were calculated based on subsets of alveolar and tumor lipid*omes* of *sub-cohort 1* [146]. Consequently, 22 tumor lipid*omes* and 21 alveolar lipd*omes* were individually combined with the respective histology scores and clinical data within a PLS regression model. Gender, age, BMI, PY, and the severity of COPD according to the GOLD were included as clinical data. Categorical variables were transferred into numeric factors before PLS regression. Furthermore, the Hb content of the tissue homogenates was associated to the lipid*omes* in the same models. With the lipid*ome* data as predictor variables and all other parameters as response variables, PLS regression models were calculated with the `plsreg2` function from the *R* package `plsdepot` [188]. Lipid species were included that were quantified in every sample. The number of included components (latent variables) into the models was automatically determined by the default cv algorithm of the `plsreg2` function. The dataset was split into ten segments. Consecutively, one of these ten segments was used for validation and the other nine for model building (leave-out-one cv). This was repeated ten times that each segment was used for validation once. Afterwards, a Q^2 was reported which was assumed to be significant when greater or equal to 0.0975 [189].

Follow-up study

In the follow-up study, the modeling strategy was optimized since a higher number of individuals was available. The complete set of available individuals was randomly split into training and validation sets with a split factor of 50 %. This accomplished that the subset that was used for

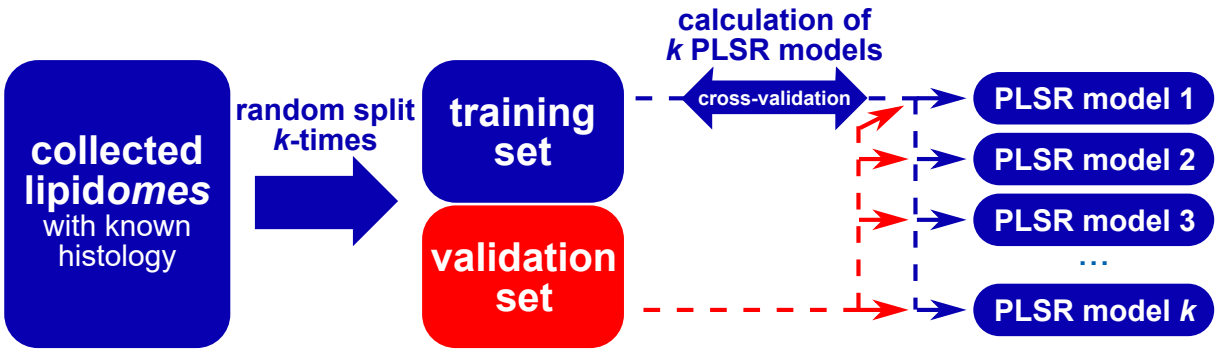


Figure 4.6: **Strategy to calculate partial least-squares regression models in the follow-up study.**

model creation was randomly composed. Overall, the datasets were split 100 times, yielding in 100 training datasets with 100 corresponding validation datasets. Lipid*omes* were used as predictor variables whereas all other supplemental parameters, such as histological characterizations and clinical data, were used as response variables. Afterwards, for each training set a PLS regression model was calculated (Fig. 4.6). PLS regressions were computed utilizing the `plsreg2` function from the *R* package `plsdepot` [188]. PLS regressions were calculated for 3 components. The cv procedure was applied to the training subset (Fig. 4.6) as described for the pilot study.

Partial least-squares regression of tumor tissue lipid*omes*. PLS regressions were calculated independently for subsets of alveolar and tumor lipid*omes*. This was necessary due to the distinct histological phenotype of both tissues. Categorical variables were transferred into numeric factors before computing PLS regressions: Gender was noted with 0 for male and 1 for female. The tumor entity was expressed by 0 for SCC, 1 for LCC, 2 for ADC, 3 for carcinoid, 4 for SCA and 5 for NET.

One tumor lipid*ome*, ID125_T, was excluded since this sample originated from a lymph node and not from a malignant tumor. All tissues without any record regarding clinical data and histological characterizations were excluded from the analysis (Tab. 4.11). Moreover, all tumor tissues with incomplete histopathological characterizations were excluded (Tab. 4.11). Lipid*ome* data from 62 tumor tissues with full histology record remained. PY values and GOLD stages were not included since these parameters were unknown for too many samples. Finally, data regarding tumor entity, necrosis, vital, tumor, stroma, inflammation, hemoglobin, BMI, gender, and age were included into the PLS regressions for tumor tissues.

Table 4.11: **Lipid*omes* excluded from partial least-squares regression.**

***Excluded:* Missing clinical and histopathological data**	***Excluded:* Incomplete histopathological data**
ID107_T, ID161_T, ID26_T, ID63_T, ID65_T	ID104_T, ID105_T, ID113_T, ID114_T, ID119_T, ID128_T, ID164_T, ID23_T, ID41_T, ID45_T, ID49_T, ID54_T, ID55_T, ID58_T, ID62_T, ID76_T, ID85_T, ID87_T, ID93_T, ID94_T
ID107_A, ID26_A, ID41_A, ID45_A, ID63_A, ID65_A, ID87_A	ID100_A, ID104_A, ID105_A, ID113_A, ID114_A, ID119_A, ID123_A, ID128_A, ID23_A, ID45_A, ID49_A, ID55_A, ID58_A, ID76_A, ID85_A, ID93_A, ID94_A

Data from 127 lipid species were included into this analysis that were present in at least 90 % of all 62 tumor samples. Missing values were imputed using the `imputePCA` function from the *R* package `MissMDA` (section 4.6.4, page 121).

With the same approach, PLS regression models were calculated for a subset of 32 ADC lipid*omes*. Furthermore, one PLS regression was calculated including all 32 ADC lipid*omes* without creation of training and validation subsets.

Partial least-squares regression of alveolar tissue lipid*omes*. Using the same approach, lipid*omes* of alveolar tissues were analyzed by PLS regression. Out of these collected 86 tissues, 7 were excluded since records regarding histopathology and clinical data were not available (Tab. 4.11). Furthermore, 17 lipid*omes* were excluded due to the fact that no histopathological characterizations were available. ID101_A was excluded since it was an outlier in the *pca* and hierarchical cluster analyses. Overall, data from 61 alveolar tissues were included into the analysis. As response variables, the histology scores inflammation, AM, fibrosis and emphysema grade were included. The patient related parameters, entity of the corresponding primary tumor, age, gender, and BMI, were included as well as the hemoglobin content of tissue homogenates. The collected parameters GOLD stage and PY could not be included into this analysis since these values were less frequently reported for the collected tissues. From this dataset, 29 patients were selected who had ADC diagnosis and were analyzed in a separate model using the same methodology.

References

[1] Y.-Y. Zhao, H. Miao, X.-L. Cheng, and F. Wei, "Lipidomics: Novel insight into the biochemical mechanism of lipid metabolism and dysregulation-associated disease," *Chemico-Biological Interactions*, vol. 240, pp. 220–238, 2015.

[2] R. Sutphen, Y. Xu, G. D. Wilbanks, J. Fiorica, E. C. Grendys, J. P. LaPolla, H. Arango, M. S. Hoffman, M. Martino, K. Wakeley, D. Griffin, R. W. Blanco, A. B. Cantor, Y.-j. Xiao, and J. P. Krischer, "Lysophospholipids are potential biomarkers of ovarian cancer," *Cancer Epidemiology Biomarkers & Prevention*, vol. 13, no. 7, pp. 1185–1191, 2004.

[3] M. Meleh, B. Pozlep, A. Mlakar, H. Meden-Vrtovec, and L. Zupancic-Kralj, "Determination of serum lysophosphatidic acid as a potential biomarker for ovarian cancer," *Journal of Chromatography B*, vol. 858, no. 1, pp. 287–291, 2007.

[4] Y. Guo, X. Wang, L. Qiu, X. Qin, H. Liu, Y. Wang, F. Li, X. Wang, G. Chen, G. Song, F. Li, S. Guo, and Z. Li, "Probing gender-specific lipid metabolites and diagnostic biomarkers for lung cancer using fourier transform ion cyclotron resonance mass spectrometry," *Clinica Chimica Acta*, vol. 414, pp. 135–141, 2012.

[5] E. D. Telenga, R. F. Hoffmann, R. t'Kindt, S. J. M. Hoonhorst, B. W. M. Willemse, A. J. M. van Oosterhout, I. H. Heijink, M. van den Berge, L. Jorge, P. Sandra, D. S. Postma, K. Sandra, and N. H. T. ten Hacken, "Untargeted lipidomic analysis in chronic obstructive pulmonary disease. Uncovering sphingolipids," *American Journal of Respiratory and Critical Care Medicine*, vol. 190, no. 2, pp. 155–164, 2014.

[6] S. Ghosh, J. C. Strum, and R. M. Bell, "Lipid biochemistry: Functions of glycerolipids and sphingolipids in cellular signaling," *The FASEB Journal*, vol. 11, no. 1, pp. 45–50, 1997.

[7] F. Perrotti, C. Rosa, I. Cicalini, P. Sacchetta, P. Del Boccio, D. Genovesi, and D. Pieragostino, "Advances in lipidomics for cancer biomarkers discovery," *International Journal of Molecular Sciences*, vol. 17, no. 12, p. 1992, 2016.

[8] J. Folch, I. Ascoli, M. Lees, J. A. Meath, and F. N. LeBaron, "Preparation of lipide extracts from brain tissue," *Journal of Biological Chemistry*, vol. 191, no. 2, pp. 833–841, 1951.

[9] E. G. Bligh and W. J. Dyer, "A rapid method of total lipid extraction and purification," *Canadian Journal of Biochemistry and Physiology*, vol. 37, no. 8, pp. 911–917, 1959.

[10] M. A. Surma, R. Herzog, A. Vasilj, C. Klose, N. Christinat, D. Morin-Rivron, K. Simons, M. Masoodi, and J. L. Sampaio, "An automated shotgun lipidomics platform for high throughput, comprehensive, and quantitative analysis of blood plasma intact lipids," *European Journal of Lipid Science and Technology*, vol. 117, no. 10, pp. 1540–1549, 2015.

[11] X. L. Guan, G. Cestra, G. Shui, A. Kuhrs, R. B. Schittenhelm, E. Hafen, F. G. van der Goot, C. C. Robinett, M. Gatti, M. Gonzalez-Gaitan, and M. R. Wenk, "Biochemical membrane lipidomics during drosophila development," *Dev Cell*, vol. 24, no. 1, pp. 98–111, 2013.

[12] K. Schuhmann, R. Herzog, D. Schwudke, W. Metelmann-Strupat, S. R. Bornstein, and A. Shevchenko, "Bottom-up shotgun lipidomics by higher energy collisional dissociation on LTQ orbitrap mass spectrometers," *Anal Chem*, vol. 83, no. 14, pp. 5480–7, 2011.

[13] X. Han and R. W. Gross, "Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: A bridge to lipidomics," *Journal of lipid research*, vol. 44, no. 6, pp. 1071–1079, 2003.

[14] D. Schwudke, K. Schuhmann, R. Herzog, S. R. Bornstein, and A. Shevchenko, "Shotgun lipidomics on high resolution mass spectrometers," *Cold Spring Harbor Perspectives in Biology*, vol. 3, no. 9, 2011.

[15] X. Han, "Multi-dimensional mass spectrometry-based shotgun lipidomics and the altered lipids at the mild cognitive impairment stage of alzheimer's disease," *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, vol. 1801, no. 8, pp. 774–783, 2010.

[16] E. Schwarz, S. Prabakaran, P. Whitfield, H. Major, F. M. Leweke, D. Koethe, P. McKenna, and S. Bahn, "High throughput lipidomic profiling of schizophrenia and bipolar disorder brain tissue reveals alterations of free fatty acids, phosphatidylcholines, and ceramides," *Journal of Proteome Research*, vol. 7, no. 10, pp. 4266–4277, 2008.

[17] J. M. Weir, G. Wong, C. K. Barlow, M. A. Greeve, A. Kowalczyk, L. Almasy, A. G. Comuzzie, M. C. Mahaney, J. B. M. Jowett, J. Shaw, J. E. Curran, J. Blangero, and P. J. Meikle, "Plasma lipid profiling in a large population-based cohort," *Journal of Lipid Research*, vol. 54, no. 10, pp. 2898–2908, 2013.

[18] K. H. Pietiläinen, M. Sysi-Aho, A. Rissanen, T. Seppänen-Laakso, H. Yki-Järvinen, J. Kaprio, and M. Orešic, "Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects - A monocytic twin study," *PLoS One*, vol. 2, no. 2, p. e218, 2007.

[19] J. M. Haus, S. R. Kashyap, T. Kasumov, R. Zhang, K. R. Kelly, R. A. DeFronzo, and J. P. Kirwan, "Plasma ceramides are elevated in obese subjects with type 2 diabetes and correlate with the severity of insulin resistance," *Diabetes*, vol. 58, no. 2, pp. 337–343, 2009.

[20] M. Ollero, G. Astarita, I. C. Guerrera, I. Sermet-Gaudelus, S. Trudel, D. Piomelli, and A. Edelman, "Plasma lipidomics reveals potential prognostic signatures within a cohort of cystic fibrosis patients," *Journal of Lipid Research*, vol. 52, no. 5, pp. 1011–1022, 2011.

[21] Y. P. Kang, W. J. Lee, J. Y. Hong, S. B. Lee, J. H. Park, D. Kim, S. Park, C.-S. Park, S.-W. Park, and S. W. Kwon, "Novel approach for analysis of bronchoalveolar lavage fluid (BALF) using HPLC-QTOF-MS-based lipidomics: Lipid levels in asthmatics and corticosteroid-treated asthmatic patients," *Journal of Proteome Research*, vol. 13, no. 9, pp. 3919–3929, 2014.

[22] Z. Hall, Z. Ament, C. H. Wilson, D. L. Burkhart, T. Ashmore, A. Koulman, T. Littlewood, G. I. Evan, and J. L. Griffin, "Myc expression drives aberrant lipid metabolism in lung cancer," *Cancer Research*, vol. 76, no. 16, pp. 4608–4618, 2016.

[23] M. Agassandian and R. K. Mallampalli, "Surfactant phospholipid metabolism," *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, vol. 1831, no. 3, pp. 612–625, 2013.

[24] F. Valianpour, J. J. M. Selhorst, L. E. M. van Lint, A. H. van Gennip, R. J. A. Wanders, and S. Kemp, "Analysis of very long-chain fatty acids using electrospray ionization mass spectrometry," *Molecular Genetics and Metabolism*, vol. 79, no. 3, pp. 189–196, 2003.

[25] H. W. Cook and C. R. McMaster, "Chapter 7 fatty acid desaturation and chain elongation in eukaryotes," *New Comprehensive Biochemistry*, vol. 36, pp. 181–204, 2002.

[26] B. Jenkins, J. West, and A. Koulman, "A review of odd-chain fatty acid metabolism and the role of pentadecanoic acid (C15:0) and heptadecanoic acid (C17:0) in health and disease," *Molecules*, vol. 20, no. 2, p. 2425, 2015.

[27] P. C. Calder, "Functional roles of fatty acids and their effects on human health," *Journal of Parenteral and Enteral Nutrition*, vol. 39, no. 1_suppl, pp. 18S–32S, 2015.

[28] "The nomenclature of lipids," *European Journal of Biochemistry*, vol. 79, no. 1, pp. 11–21, 1977.

[29] M. Prentki and S. R. M. Madiraju, "Glycerolipid/free fatty acid cycle and islet β-cell function in health, obesity and diabetes," *Molecular and Cellular Endocrinology*, vol. 353, no. 1, pp. 88–100, 2012.

[30] D. Schwudke, J. T. Hannich, V. Surendranath, V. Grimard, T. Moehring, L. Burton, T. Kurzchalia, and A. Shevchenko, "Top-down lipidomic screens by multivariate analysis of high-resolution survey mass spectra," *Analytical Chemistry*, vol. 79, no. 11, pp. 4083–4093, 2007.

[31] C. Kent, "Eukaryotic phospholipid biosynthesis," *Annual Review of Biochemistry*, vol. 64, no. 1, pp. 315–343, 1995.

[32] K.-A. Karlsson, "On the chemistry and occurrence of sphingolipid long-chain bases," *Chemistry and Physics of Lipids*, vol. 5, no. 1, pp. 6–43, 1970.

[33] A. H. Merrill, "De novo sphingolipid biosynthesis: A necessary, but dangerous, pathway," *Journal of Biological Chemistry*, vol. 277, no. 29, pp. 25843–25846, 2002.

[34] Y. Chen, Y. Liu, M. C. Sullards, and A. H. Merrill, "An introduction to sphingolipid metabolism and analysis by new technologies," *NeuroMolecular Medicine*, vol. 12, no. 4, pp. 306–319, 2010.

[35] N. Bartke and Y. A. Hannun, "Bioactive sphingolipids: metabolism and function," *Journal of Lipid Research*, vol. 50, no. Supplement, pp. S91–S96, 2009.

[36] D. W. Russell, "Cholesterol biosynthesis and metabolism," *Cardiovascular Drugs and Therapy*, vol. 6, no. 2, pp. 103–110, 1992.

[37] H. Tapiero, G. Nguyen Ba, P. Couvreur, and K. D. Tew, "Polyunsaturated fatty acids (PUFA) and eicosanoids in human health and pathologies," *Biomedicine and Pharmacotherapy*, vol. 56, no. 5, pp. 215–222, 2002.

[38] S. Martin and R. G. Parton, "Lipid droplets: A unified view of a dynamic organelle," *Nature Reviews. Molecular Cell Biology*, vol. 7, no. 5, pp. 373–378, 2006.

[39] S. Carrasco and I. Mérida, "Diacylglycerol, when simplicity becomes complex," *Trends in Biochemical Sciences*, vol. 32, no. 1, pp. 27–36, 2007.

[40] M. Chabanon, J. C. Stachowiak, and P. Rangamani, "Systems biology of cellular membranes: a convergence with biophysics," *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, vol. 9, no. 5, pp. e1386–n/a, 2017.

[41] J. G. Alb Jr, M. A. Kearns, and V. A. Bankaitis, "Phospholipid metabolism and membrane dynamics," *Current Opinion in Cell Biology*, vol. 8, no. 4, pp. 534–541, 1996.

[42] G. van Meer, D. R. Voelker, and G. W. Feigenson, "Membrane lipids: Where they are and how they behave," *Nature Reviews. Molecular Cell Biology*, vol. 9, no. 2, pp. 112–124, 2008.

[43] D. Hishikawa, T. Hashidate, T. Shimizu, and H. Shindou, "Diversity and function of membrane glycerophospholipids generated by the remodeling pathway in mammalian cells," *Journal of Lipid Research*, vol. 55, no. 5, pp. 799–807, 2014.

[44] E. L. Crockett, "Cholesterol function in plasma membranes from ectotherms: Membrane-specific roles in adaptation to temperature," *American Zoologist*, vol. 38, no. 2, pp. 291–304, 1998.

[45] N. M. F. S. A. Cerqueira, E. F. Oliveira, D. S. Gesto, D. Santos-Martins, C. Moreira, H. N. Moorthy, M. J. Ramos, and P. A. Fernandes, "Cholesterol biosynthesis: A mechanistic overview," *Biochemistry*, vol. 55, no. 39, pp. 5483–5506, 2016.

[46] K. Simons and E. Ikonen, "Functional rafts in cell membranes," *Nature*, vol. 387, no. 6633, pp. 569–572, 1997.

[47] Y. A. Hannun and R. M. Bell, "Functions of sphingolipids and sphingolipid breakdown products in cellular regulation," *Science*, vol. 243, no. 4890, pp. 500–507, 1989.

[48] E. Rodriguez-Boulan and W. Nelson, "Morphogenesis of the polarized epithelial cell phenotype," *Science*, vol. 245, no. 4919, pp. 718–725, 1989.

[49] *National Heart, Lung and Blood Institute, The Respiratory System*, 2017. https://www.nhlbi.nih.gov/health/health-topics/topics/hlw/system.

[50] K. Chen and J. K. Kolls, “Good and bad lipids in the lung,” *Nature Medicine*, vol. 16, no. 10, pp. 1078–1079, 2010.

[51] E. Schneeberger, *Alveolar type I cells.* In The Lung Scientific Foundations, Philadelphia: Lippincott-Raven Publishers, 2 ed., 1997.

[52] L. Guillot, N. Nathan, O. Tabary, G. Thouvenin, P. Le Rouzic, H. Corvol, S. Amselem, and A. Clement, “Alveolar epithelial cells: Master regulators of lung homeostasis,” *The International Journal of Biochemistry and Cell Biology*, vol. 45, no. 11, pp. 2568–2573, 2013.

[53] R. J. Mason, “Biology of alveolar type II cells,” *Respirology*, vol. 11, pp. S12–S15, 2006.

[54] V. Goss, A. N. Hunt, and A. D. Postle, “Regulation of lung surfactant phospholipid synthesis and metabolism,” *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, vol. 1831, no. 2, pp. 448–458, 2013.

[55] A. D. Postle, L. W. Gonzales, W. Bernhard, G. T. Clark, M. H. Godinez, R. I. Godinez, and P. L. Ballard, “Lipidomics of cellular and secreted phospholipids from differentiated human fetal type II alveolar epithelial cells,” *Journal of Lipid Research*, vol. 47, no. 6, pp. 1322–1331, 2006.

[56] C. J. Lang, A. D. Postle, S. Orgeig, F. Possmayer, W. Bernhard, A. K. Panda, K. D. Jürgens, W. K. Milsom, K. Nag, and C. B. Daniels, “Dipalmitoylphosphatidylcholine is not the major surfactant phospholipid species in all mammals,” *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, vol. 289, no. 5, pp. R1426–R1439, 2005.

[57] K. Kuronuma, H. Mitsuzawa, K. Takeda, C. Nishitani, E. D. Chan, Y. Kuroki, M. Nakamura, and D. R. Voelker, “Anionic pulmonary surfactant phospholipids inhibit inflammatory responses from alveolar macrophages and U937 cells by binding the lipopolysaccharide-interacting proteins CD14 and MD-2,” *Journal of Biological Chemistry*, vol. 284, no. 38, pp. 25488–25500, 2009.

[58] C. B. Daniels, H. A. Barr, J. H. Power, and T. E. Nicholas, “Body temperature alters the lipid composition of pulmonary surfactant in the lizard ctenophorus nuchalis,” *Experimental lung research*, vol. 16, no. 5, pp. 435–449, 1990.

[59] A. D. Postle, E. L. Heeley, and D. C. Wilton, “A comparison of the molecular species compositions of mammalian lung surfactant phospholipids,” *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, vol. 129, no. 1, pp. 65–73, 2001.

[60] M. Numata, H. W. Chu, A. Dakhama, and D. R. Voelker, “Pulmonary surfactant phosphatidylglycerol inhibits respiratory syncytial virus-induced inflammation and infection,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 1, pp. 320–325, 2010.

[61] M. Hallman, G. Enhorning, and F. Possmayer, “Composition and surface activity of normal and phosphatidylglycerol-deficient lung surfactant,” *Pediatric Research*, vol. 19, no. 3, pp. 286–292, 1985.

[62] M. Griese, H. G. Kirmeier, G. Liebisch, D. Rauch, F. Stückler, G. Schmitz, R. Zarbock, and ILD-BAL working group of the Kids-Lung-Register, “Surfactant lipidomics in healthy children and childhood interstitial lung disease,” *PLoS One*, vol. 10, no. 2, p. e0117985, 2015.

[63] N. B. Ray, L. Durairaj, B. B. Chen, B. J. McVerry, A. J. Ryan, M. Donahoe, A. K. Waltenbaugh, C. P. O’Donnell, F. C. Henderson, C. A. Etscheidt, D. M. McCoy, M. Agassandian, E. C. Hayes-Rowan, T. A. Coon, P. L. Butler, L. Gakhar, S. N. Mathur, J. C. Sieren, Y. Y. Tyurina, V. E. Kagan, G. McLennan, and R. K. Mallampalli, “Dynamic regulation of cardiolipin by the lipid pump Atp8b1 determines the severity of lung injury in experimental pneumonia,” *Nature Medicine*, vol. 16, no. 10, pp. 1120–1127, 2010.

[64] I. Petrache, V. Natarajan, L. Zhen, T. R. Medler, A. Richter, C. Cho, W. C. Hubbard, E. V. Berdyshev, and R. M. Tuder, “Ceramide upregulation causes pulmonary cell apoptosis and emphysema,” *Nature medicine*, vol. 11, no. 5, pp. 491–498, 2005.

[65] A. P. Seitz, H. Grassmé, M. J. Edwards, Y. Pewzner-Jung, and E. Gulbins, “Ceramide and sphingosine in pulmonary infections,” *Biological Chemistry*, vol. 396, no. 6-7, p. 611, 2015.

[66] H. S. Kang, S. C. Lee, Y. S. Park, Y. E. Jeon, J. H. Lee, S.-Y. Jung, I. H. Park, S. H. Jang, H. M. Park, C. W. Yoo, S. H. Park, S. Y. Han, K. P. Kim, Y. H. Kim, J. Ro, and H. K. Kim, “Protein and lipid MALDI profiles classify breast cancers according to the intrinsic subtype,” *BMC Cancer*, vol. 11, no. 1, p. 465, 2011.

[67] M. Hilvo, C. Denkert, L. Lehtinen, B. Müller, S. Brockmöller, T. Seppänen-Laakso, J. Budczies, E. Bucher, L. Yetukuri, S. Castillo, E. Berg, H. Nygren, M. Sysi-Aho, J. L. Griffin, O. Fiehn, S. Loibl, C. Richter-Ehrenstein, C. Radke, T. Hyötyläinen, O. Kallioniemi, K. Iljin, and M. Orešic, “Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression,” *Cancer Research*, vol. 71, no. 9, pp. 3236–3245, 2011.

[68] E. Cífkoá, M. Holcapek, M. Lísa, D. Vrána, B. Melichar, and V. Študent, “Lipidomic differentiation between human kidney tumors and surrounding normal tissues using HILIC-HPLC/ESI-MS and multivariate data analysis,” *Journal of Chromatography B*, vol. 1000, pp. 14–21, 2015.

[69] T. Goto, N. Terada, T. Inoue, K. Nakayama, Y. Okada, T. Yoshikawa, Y. Miyazaki, M. Uegaki, S. Sumiyoshi, T. Kobayashi, T. Kamba, K. Yoshimura, and O. Ogawa, “The expression profile of phosphatidylinositol in high spatial resolution imaging mass spectrometry as a potential biomarker for prostate cancer,” *PLoS One*, vol. 9, no. 2, p. e90242, 2014.

[70] T. Goto, N. Terada, T. Inoue, T. Kobayashi, K. Nakayama, Y. Okada, T. Yoshikawa, Y. Miyazaki, M. Uegaki, N. Utsunomiya, Y. Makino, S. Sumiyoshi, T. Yamasaki, T. Kamba, and O. Ogawa, “Decreased expression of lysophosphatidylcholine (16:0/OH) in high resolution imaging mass spectrometry independently predicts biochemical recurrence after surgical treatment for prostate cancer,” *Prostate*, vol. 75, no. 16, pp. 1821–30, 2015.

[71] I. Dobrzyńska, B. Szachowicz-Petelska, S. Sulkowski, and Z. Figaszewski, “Changes in electric charge and phospholipids composition in human colorectal cancer cells,” *Molecular and Cellular Biochemistry*, vol. 276, no. 1, pp. 113–119, 2005.

[72] N. Kurabe, T. Hayasaka, M. Ogawa, N. Masaki, Y. Ide, M. Waki, T. Nakamura, K. Kurachi, T. Kahyo, K. Shinmura, Y. Midorikawa, Y. Sugiyama, M. Setou, and H. Sugimura, “Accumulated phosphatidylcholine (16:0/16:1) in human colorectal cancer; possible involvement of LPCAT4,” *Cancer Sci*, vol. 104, no. 10, pp. 1295–302, 2013.

[73] S. Kang, A. Lee, Y. S. Park, S. C. Lee, S. Y. Park, S. Y. Han, K. P. Kim, Y. H. Kim, C. W. Yoo, and H. K. Kim, “Alteration in lipid and protein profiles of ovarian cancer: Similarity to breast cancer,” *International Journal of Gynecological Cancer*, vol. 21, no. 9, pp. 1566–1572, 2011.

[74] G. K. Lee, H. S. Lee, Y. S. Park, J. H. Lee, S. C. Lee, J. H. Lee, S. J. Lee, S. R. Shanta, H. M. Park, H. R. Kim, I. H. Kim, Y. H. Kim, J. I. Zo, K. P. Kim, and H. K. Kim, “Lipid MALDI profile classifies non-small cell lung cancers according to the histologic type,” *Lung Cancer*, vol. 76, no. 2, pp. 197–203, 2012.

[75] E. Marien, M. Meister, T. Muley, S. Fieuws, S. Bordel, R. Derua, J. Spraggins, R. Van de Plas, J. Dehairs, J. Wouters, M. Bagadi, H. Dienemann, M. Thomas, P. A. Schnabel, R. M. Caprioli, E. Waelkens, and J. V. Swinnen, “Non-small cell lung cancer is characterized by dramatic changes in phospholipid profiles,” *International Journal of Cancer*, vol. 137, no. 7, pp. 1539–1548, 2015.

[76] K. Chughtai, L. Jiang, T. R. Greenwood, K. Glunde, and R. M. A. Heeren, “Mass spectrometry images acylcarnitines, phosphatidylcholines, and sphingomyelin in MDA-MB-231 breast tumor models,” *Journal of Lipid Research*, vol. 54, no. 2, pp. 333–344, 2013.

[77] S. Guo, L. Qiu, Y. Wang, X. Qin, H. Liu, M. He, Y. Zhang, Z. Li, and X. Chen, “Tissue imaging and serum lipidomic profiling for screening potential biomarkers of thyroid tumors by matrix-assisted laser desorption/ionization-fourier transform ion cyclotron resonance mass spectrometry,” *Analytical and Bioanalytical Chemistry*, vol. 406, no. 18, pp. 4357–4370, 2014.

[78] T. Uehara, H. Kikuchi, S. Miyazaki, I. Iino, T. Setoguchi, Y. Hiramatsu, M. Ohta, K. Kamiya, Y. Morita, H. Tanaka, S. Baba, T. Hayasaka, M. Setou, and H. Konno, “Overexpression of lysophosphatidylcholine acyltransferase 1 and concomitant lipid alterations in gastric cancer,” *Annals of Surgical Oncology*, vol. 23, no. 2, pp. 206–213, 2016.

[79] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2015,” *CA: A Cancer Journal for Clinicians*, vol. 65, no. 1, pp. 5–29, 2015.

[80] A. J. Alberg and J. M. Samet, “Epidemiology of lung cancer,” *Chest*, vol. 123, no. 1_suppl, pp. 21S–49S, 2003.

[81] J. Subramanian and R. Govindan, “Lung cancer in never smokers: A review,” *Journal of Clinical Oncology*, vol. 25, no. 5, pp. 561–570, 2007.

[82] B. Phillips, E. Veljkovic, S. Boué, W. K. Schlage, G. Vuillaume, F. Martin, B. Titz, P. Leroy, A. Buettner, A. Elamin, A. Oviedo, M. Cabanski, H. De León, E. Guedj, T. Schneider, M. Talikka, N. V. Ivanov, P. Vanscheeuwijck, M. C. Peitsch, and J. Hoeng, "An 8-month systems toxicology inhalation/cessation study in Apoe(-/-) mice to investigate cardiovascular and respiratory exposure effects of a candidate modified risk tobacco product, THS 2.2, compared with conventional cigarettes," *Toxicological Sciences*, vol. 149, no. 2, pp. 411–432, 2016.

[83] B. Titz, S. Boué, B. Phillips, M. Talikka, T. Vihervaara, T. Schneider, C. Nury, A. Elamin, E. Guedj, and M. J. Peck, "Effects of cigarette smoke, cessation, and switching to two heat-not-burn tobacco products on lung lipid metabolism in C57BL/6 and Apoe-/- mice - an integrative systems toxicology analysis," *Toxicological Sciences*, vol. 149, no. 2, pp. 441–457, 2016.

[84] E. Marien, M. Meister, T. Muley, T. G. d. Pulgar, R. Derua, J. M. Spraggins, R. V. d. Plas, F. Vanderhoydonc, J. Machiels, M. M. Binda, J. Dehairs, J. Willette-Brown, Y. Hu, H. Dienemann, M. Thomas, P. A. Schnabel, R. M. Caprioli, J. C. Lacal, E. Waelkens, and J. V. Swinnen, "Phospholipid profiling identifies acyl chain elongation as a ubiquitous trait and potential target for the treatment of lung squamous cell carcinoma," *Oncotarget*, vol. 7, no. 11, pp. 12582–12597, 2016.

[85] T. A. Stamey, N. Yang, A. R. Hay, J. E. McNeal, F. S. Freiha, and E. Redwine, "Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate," *New England Journal of Medicine*, vol. 317, no. 15, pp. 909–916, 1987. PMID: 2442609.

[86] Y. Zhang, Y. Liu, L. Li, J. Wei, S. Xiong, and Z. Zhao, "High resolution mass spectrometry coupled with multivariate data analysis revealing plasma lipidomic alteration in ovarian cancer in asian women," *Talanta*, vol. 150, pp. 88–96, 2016.

[87] Z. Zhao, Y. Xiao, P. Elson, H. Tan, S. J. Plummer, M. Berk, P. P. Aung, I. C. Lavery, J. P. Achkar, L. Li, G. Casey, and Y. Xu, "Plasma lysophosphatidylcholine levels: Potential biomarkers for colorectal cancer," *Journal of Clinical Oncology*, vol. 25, no. 19, pp. 2696–2701, 2007.

[88] H.-Y. Kim, K.-M. Lee, S.-H. Kim, Y.-J. Kwon, Y.-J. Chun, and H.-K. Choi, "Comparative metabolic and lipidomic profiling of human breast cancer cells with different metastatic potentials," *Oncotarget*, vol. 7, no. 41, pp. 67111–67128, 2016.

[89] S. Sales, J. Graessler, S. Ciucci, R. Al-Atrib, T. Vihervaara, K. Schuhmann, D. Kauhanen, M. Sysi-Aho, S. R. Bornstein, M. Bickle, C. V. Cannistraci, K. Ekroos, and A. Shevchenko, "Gender, contraceptives and individual metabolic predisposition shape a healthy plasma lipidome," *Scientific Reports*, vol. 6, p. 27710, 2016.

[90] J. Olert, K.-H. Wiedorn, T. Goldmann, H. Kühl, Y. Mehraein, H. Scherthan, F. Niketeghad, E. Vollmer, A. M. Müller, and J. Müller-Navia, "HOPE fixation: A novel fixing method and paraffin-embedding technique for human soft tissues," *Pathology - Research and Practice*, vol. 197, no. 12, pp. 823–826, 2001.

[91] A. H. Fischer, K. A. Jacobson, J. Rose, and R. Zeller, "Hematoxylin and eosin staining of tissue and cell sections," *Cold Spring Harbor Protocols*, vol. 2008, no. 5, p. pdb.prot4986, 2008.

[92] R. D. Cardiff, C. H. Miller, and R. J. Munn, "Manual hematoxylin and eosin staining of mouse tissue sections," *Cold Spring Harbor Protocols*, vol. 2014, no. 6, p. pdb.prot073411, 2014.

[93] M. C. Lock, E. V. McGillick, S. Orgeig, S. Zhang, I. C. McMillen, and J. L. Morrison, "Mature surfactant protein-B expression by immunohistochemistry as a marker for surfactant system development in the fetal sheep lung," *Journal of Histochemistry & Cytochemistry*, vol. 63, no. 11, pp. 866–878, 2015.

[94] D. S. Phelps and J. Floros, "Localization of pulmonary surfactant proteins using immunohistochemistry and tissue in situ hybridization," *Experimental Lung Research*, vol. 17, no. 6, pp. 985–995, 1991.

[95] L. Maneta-Peyret, P. Compére, P. Moreau, G. Goffinet, and C. Cassagne, "Immunocytochemistry of lipids: Chemical fixatives have dramatic effects on the preservation of tissue lipids," *The Histochemical Journal*, vol. 31, no. 8, pp. 541–547, 1999.

[96] K. Arafah, R. Longuespée, A. Desmons, O. Kerdraon, I. Fournier, and M. Salzet, "Lipidomics for clinical diagnosis: Dye-assisted laser desorption/ionization (DALDI) method for lipids detection in MALDI mass spectrometry imaging," *Omics: a journal of integrative biology*, vol. 18, no. 8, pp. 487–498, 2014.

[97] K. L. Burdon, "Fatty material in bacteria and fungi revealed by staining dried, fixed slide preparations," *Journal of Bacteriology*, vol. 52, no. 6, pp. 665–678, 1946.

[98] J. A. Considine and R. B. Knox, "Development and histochemistry of the cells, cell walls, and cuticle of the dermal system of fruit of the grape,vitis vinifera l," *Protoplasma*, vol. 99, no. 4, pp. 347–365, 1979.

[99] J. L. Ramírez-Zacarías, F. Castro-Munozledo, and W. Kuri-Harcuch, "Quantitation of adipose conversion and triglycerides by staining intracytoplasmic lipids with oil red O," *Histochemistry*, vol. 97, no. 6, pp. 493–497, 1992.

[100] A. Römpp and B. Spengler, "Mass spectrometry imaging with high resolution in mass and space," *Histochemistry and Cell Biology*, vol. 139, no. 6, pp. 759–783, 2013.

[101] M. Karas, D. Bachmann, U. Bahr, and F. Hillenkamp, "Matrix-assisted ultraviolet laser desorption of non-volatile compounds," *International Journal of Mass Spectrometry and Ion Processes*, vol. 78, pp. 53–68, 1987.

[102] L. F. Eggers and D. Schwudke, *Lipid Extraction: Basics of the Methyl-tert-Butyl Ether Extraction*, pp. 1–3. Dordrecht: Springer Netherlands, 2016.

[103] W. R. Bloor, "The determination of small amounts of lipid in blood plasma," *Journal of Biological Chemistry*, vol. 77, no. 1, pp. 53–73, 1928.

[104] J. Folch, M. Lees, and G. H. S. Stanley, "A simple method for the isolation and purification of total lipides from animal tissues," *Journal of Biological Chemistry*, vol. 226, no. 1, pp. 497–509, 1957.

[105] V. Matyash, G. Liebisch, T. V. Kurzchalia, A. Shevchenko, and D. Schwudke, “Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics,” *Journal of Lipid Research*, vol. 49, no. 5, pp. 1137–1146, 2008.

[106] R. v. Noorden, B. Maher, and R. Nuzzo, “The top 100 papers,” *Nature*, vol. 514, no. 7524, pp. 550–553, 2014.

[107] L. F. Eggers and D. Schwudke, *Liquid Extraction: Folch*, pp. 1–6. Dordrecht: Springer Netherlands, 2016.

[108] A. Sündermann, L. F. Eggers, and D. Schwudke, *Liquid Extraction: Bligh and Dyer*, pp. 1–4. Dordrecht: Springer Netherlands, 2016.

[109] S. Chen, M. Hoene, J. Li, Y. Li, X. Zhao, H.-U. Häring, E. D. Schleicher, C. Weigert, G. Xu, and R. Lehmann, “Simultaneous extraction of metabolome and lipidome with methyl tert-butyl ether from a single small tissue sample for ultra-high performance liquid chromatography/mass spectrometry,” *Journal of Chromatography A*, vol. 1298, pp. 9–16, 2013.

[110] W. Nernst, “Über die Verteilung eines Stoffes zwischen zwei Lösungsmitteln,” *Nachrichten von der Königl. Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen*, pp. 401–416, 1890.

[111] T. Mehmood and B. Ahmed, “The diversity in the applications of partial least squares: An overview,” *Journal of Chemometrics*, vol. 30, no. 1, pp. 4–17, 2016.

[112] H. K. Mangold, “Thin-layer chromatography of lipids,” *Journal of the American Oil Chemists Society*, vol. 38, no. 12, pp. 708–727, 1961.

[113] J. Jouhet, J. Lupette, O. Clerc, L. Magneschi, M. Bedhomme, S. Collin, S. Roy, E. Maréchal, and F. Rébeillé, “LC-MS/MS versus TLC plus GC methods: Consistency of glycerolipid and fatty acid profiles in microalgae and higher plant cells and effect of a nitrogen starvation,” *PLoS One*, vol. 12, no. 8, p. e0182423, 2017.

[114] M. Yamashita and J. B. Fenn, “Electrospray ion source. Another variation on the free-jet theme,” *The Journal of Physical Chemistry*, vol. 88, no. 20, pp. 4451–4459, 1984.

[115] M. Yamashita and J. B. Fenn, “Negative ion production with the electrospray ion source,” *The Journal of Physical Chemistry*, vol. 88, no. 20, pp. 4671–4675, 1984.

[116] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, “Electrospray ionization - principles and practice,” *Mass Spectrometry Reviews*, vol. 9, no. 1, pp. 37–70, 1990.

[117] M. Karas, U. Bahr, and T. Dülcks, “Nano-electrospray ionization mass spectrometry: Addressing analytical problems beyond routine,” *Fresenius’ Journal of Analytical Chemistry*, vol. 366, no. 6, pp. 669–676, 2000.

[118] L. Rayleigh, “On the equilibrium of liquid conducting masses charged with electricity,” *Philosophical Magazine*, vol. 14, no. 87, pp. 184–186, 1882.

[119] M. Wilm and M. Mann, “Analytical properties of the nanoelectrospray ion source,” *Analytical chemistry*, vol. 68, no. 1, pp. 1–8, 1996.

[120] G. A. Schultz, T. N. Corso, S. J. Prosser, and S. Zhang, “A fully integrated monolithic microchip electrospray device for mass spectrometry,” *Analytical Chemistry*, vol. 72, no. 17, pp. 4058–4063, 2000.

[121] D. Schwudke, G. Liebisch, R. Herzog, G. Schmitz, and A. Shevchenko, “Shotgun lipidomics by tandem mass spectrometry under data dependent acquisition control,” *Methods in enzymology*, vol. 433, pp. 175–191, 2007.

[122] K. Schuhmann, R. Almeida, M. Baumert, R. Herzog, S. R. Bornstein, and A. Shevchenko, “Shotgun lipidomics on a LTQ orbitrap mass spectrometer by successive switching between acquisition polarity modes,” *Journal of Mass Spectrometry*, vol. 47, no. 1, pp. 96–104, 2012.

[123] B. Brügger, G. Erben, R. Sandhoff, F. T. Wieland, and W. D. Lehmann, “Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry,” *Proceedings of the National academy of Sciences*, vol. 94, no. 6, pp. 2339–2344, 1997.

[124] Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman, and R. G. Cooks, “The orbitrap: A new mass spectrometer,” *Journal of Mass Spectrometry*, vol. 40, no. 4, pp. 430–443, 2005.

[125] A. G. Marshall, C. L. Hendrickson, and G. S. Jackson, “Fourier transform ion cyclotron resonance mass spectrometry: A primer,” *Mass spectrometry reviews*, vol. 17, no. 1, pp. 1–35, 1998.

[126] A. Taichrib, M. Pelzing, C. Pellegrino, M. Rossi, and C. Neusüß, “High resolution TOF MS coupled to CE for the analysis of isotopically resolved intact proteins,” *Journal of Proteomics*, vol. 74, no. 7, pp. 958 – 966, 2011.

[127] J. Graessler, D. Schwudke, P. E. H. Schwarz, R. Herzog, A. Shevchenko, and S. R. Bornstein, “Top-down lipidomics reveals ether lipid deficiency in blood plasma of hypertensive patients,” *PLoS One*, vol. 4, no. 7, p. e6261, 2009.

[128] M. Lankinen, U. Schwab, A. Erkkilä, T. Seppänen-Laakso, M.-L. Hannila, H. Mussalo, S. Lehto, M. Uusitupa, H. Gylling, and M. Orešic, “Fatty fish intake decreases lipids related to inflammation and insulin signaling - A lipidomics approach,” *PLoS One*, vol. 4, no. 4, p. e5258, 2009.

[129] J. Li, S. Ren, H.-l. Piao, F. Wang, P. Yin, C. Xu, X. Lu, G. Ye, Y. Shao, M. Yan, X. Zhao, Y. Sun, and G. Xu, “Integration of lipidomics and transcriptomics unravels aberrant lipid metabolism and defines cholesteryl oleate as potential biomarker of prostate cancer,” *Scientific Reports*, vol. 6, p. 20984, 2016.

[130] E. L. Pannkuk, E. C. Laiakis, T. D. Mak, G. Astarita, S. Authier, K. Wong, and A. J. Fornace, “A lipidomic and metabolomic serum signature from nonhuman primates exposed to ionizing radiation,” *Metabolomics*, vol. 12, no. 5, p. 80, 2016.

[131] F. Domínguez, M. Ferrando, P. Díaz-Gimeno, F. Quintana, G. Fernández, I. Castells, and C. Simón, “Lipidomic profiling of endometrial fluid in women with ovarian endometriosis,” *Biology of Reproduction*, vol. 96, no. 4, pp. 772–779, 2017.

[132] H. W. Thompson, R. Mera, and C. Prasad, “A description of the appropriate use of student’s t-test,” *Nutritional Neuroscience*, vol. 1, no. 2, pp. 165–172, 1998.

[133] P.-C. Bürkner, P. Doebler, and H. Holling, “Optimal design of the Wilcoxon-Mann-Whitney-test,” *Biometrical Journal*, vol. 59, no. 1, pp. 25–40, 2016.

[134] S. G. Hilsenbeck and G. M. Clark, “Practical p-value adjustment for optimally selected cutpoints,” *Statistics in Medicine*, vol. 15, no. 1, pp. 103–112, 1996.

[135] Y. Benjamini and Y. Hochberg, “Controlling the false discovery rate: A practical and powerful approach to multiple testing,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[136] T. Vihervaara, M. Suoniemi, and R. Laaksonen, “Lipidomics in drug discovery,” *Drug Discovery Today*, vol. 19, no. 2, pp. 164–170, 2014.

[137] A. Vargha and H. D. Delaney, “The Kruskal-Wallis test and stochastic homogeneity,” *Journal of Educational and Behavioral Statistics*, vol. 23, no. 2, pp. 170–192, 1998.

[138] N. Gehlenborg, S. I. O’Donoghue, N. S. Baliga, A. Goesmann, M. A. Hibbs, H. Kitano, O. Kohlbacher, H. Neuweger, R. Schneider, and D. Tenenbaum, “Visualization of omics data for systems biology,” *Nature methods*, vol. 7, pp. S56–S68, 2010.

[139] P. S. Niemelä, S. Castillo, M. Sysi-Aho, and M. Orešic, “Bioinformatics and computational methods for lipidomics,” *Journal of Chromatography B*, vol. 877, no. 26, pp. 2855–2862, 2009.

[140] L. Yetukuri, M. Katajamaa, G. Medina-Gomez, T. Seppänen-Laakso, A. Vidal-Puig, and M. Orešic, “Bioinformatics strategies for lipidomics analysis: Characterization of obesity related hepatic steatosis,” *BMC Systems Biology*, vol. 1, no. 1, p. 12, 2007.

[141] P. Geladi and B. R. Kowalski, “Partial least-squares regression: A tutorial,” *Analytica Chimica Acta*, vol. 185, pp. 1–17, 1986.

[142] S. Wold, M. Sjöström, and L. Eriksson, “PLS-regression: A basic tool of chemometrics,” *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.

[143] X. Li, R. Gill, N. G. Cooper, J. K. Yoo, and S. Datta, “Modeling microRNA-mRNA interactions using PLS regression in human colon cancer,” *BMC Medical Genomics*, vol. 4, no. 1, pp. 1–15, 2011.

[144] J. Josse and F. Husson, “missMDA: A package for handling missing values in multivariate data analysis,” *2016*, vol. 70, no. 1, p. 31, 2016.

[145] C. Marella, A. E. Torda, and D. Schwudke, “The LUX score: A metric for lipidome homology,” *PLoS Computational Biology*, vol. 11, no. 9, p. e1004511, 2015.

[146] L. F. Eggers, J. Müller, C. Marella, V. Scholz, H. Watz, C. Kugler, K. F. Rabe, T. Goldmann, and D. Schwudke, "Lipidomes of lung cancer and tumour-free lung tissues reveal distinct molecular signatures for cancer differentiation, age, inflammation, and pulmonary emphysema," *Scientific Reports*, vol. 7, no. 1, p. 11087, 2017.

[147] K. F. Rabe, S. Hurd, A. Anzueto, P. J. Barnes, S. A. Buist, P. Calverley, Y. Fukuchi, C. Jenkins, R. Rodriguez-Roisin, C. van Weel, J. Zielinski, and Global Initiative for Chronic Obstructive Lung Disease, "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary," *American Journal of Respiratory and Critical Care Medicine*, vol. 176, no. 6, pp. 532–55, 2007.

[148] A. Nagai, I. Yamawaki, W. M. Thurlbeck, and T. Takizawa, "Assessment of lung parenchymal destruction by using routine histologic tissue sections," *American Review of Respiratory Disease*, vol. 139, no. 2, pp. 313–319, 1989.

[149] K. Ito and P. J. Barnes, "COPD as a disease of accelerated lung aging," *Chest*, vol. 135, no. 1, pp. 173–180, 2009.

[150] M. Nazari and D. C. Muddiman, "Enhanced lipidome coverage in shotgun analyses by using gas-phase fractionation," *Journal of The American Society for Mass Spectrometry*, pp. 1–10, 2016.

[151] G. Liebisch, K. Ekroos, M. Hermansson, and C. S. Ejsing, "Reporting of lipidomics data should be standardized," *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, vol. 1862, no. 8, pp. 747–751, 2017.

[152] E. Rampler, C. Coman, G. Hermann, A. Sickmann, R. Ahrends, and G. Koellensperger, "LILY-lipidome isotope labeling of yeast: In vivo synthesis of 13C labeled reference lipids for quantification by mass spectrometry," *Analyst*, vol. 142, pp. 1891–1899, 2017.

[153] T. Goldmann, D. Kähler, H. Schultz, M. Abdullah, D. S. Lang, F. Stellmacher, and E. Vollmer, "On the significance of Surfactant Protein-A within the human lungs," *Diagnostic Pathology*, vol. 4, p. 8, Mar 2009.

[154] C. Betz, T. Papadopoulos, J. Buchwald, J. Dämmrich, and H. K. Müller-Hermelink, "Surfactant protein gene expression in metastatic and micrometastatic pulmonary adenocarcinomas and other non-small cell lung carcinomas: Detection by reverse transcriptase-polymerase chain reaction," *Cancer Research*, vol. 55, no. 19, pp. 4283–4286, 1995.

[155] N. Zehethofer, S. Bermbach, S. Hagner, H. Garn, J. Müller, T. Goldmann, B. Lindner, D. Schwudke, and P. König, "Lipid analysis of airway epithelial cells for studying respiratory diseases," *Chromatographia*, vol. 78, no. 5, pp. 403–413, 2015.

[156] S. Hebbar, W. D. Schulz, U. Sauer, and D. Schwudke, "Laser capture microdissection coupled with on-column extraction LC-MSn enables lipidomics of fluorescently labeled drosophila neurons," *Analytical Chemistry*, vol. 86, no. 11, pp. 5345–5352, 2014.

[157] M. Kwiatkowski, M. Wurlitzer, A. Krutilin, P. Kiani, R. Nimer, M. Omidi, A. Mannaa, T. Bussmann, K. Bartkowiak, S. Kruber, S. Uschold, P. Steffen, J. Lübberstedt, N. Küpker, H. Petersen, R. Knecht, N. O. Hansen, A. Zarrine-Afsar, W. D. Robertson, R. J. D. Miller, and H. Schlüter, "Homogenization of tissues via picosecond-infrared laser (PIRL) ablation: Giving a closer view on the in-vivo composition of protein species as compared to mechanical homogenization," *Journal of Proteomics*, vol. 134, pp. 193–202, 2016.

[158] G. Pascual, A. Avgustinova, S. Mejetta, M. Martín, A. Castellanos, C. S.-O. Attolini, A. Berenguer, N. Prats, A. Toll, J. A. Hueto, C. Bescós, L. Di Croce, and S. A. Benitah, "Targeting metastasis-initiating cells through the fatty acid receptor CD36," *Nature*, vol. 541, no. 7635, pp. 41–45, 2017.

[159] F.-F. Hsu and J. Turk, "Studies on phosphatidylglycerol with triple quadrupole tandem mass spectrometry with electrospray ionization: fragmentation processes and structural characterization," *Journal of the American Society for Mass Spectrometry*, vol. 12, no. 9, pp. 1036–1043, 2001.

[160] C. E. Dodd, C. J. Pyle, R. Glowinski, M. V. S. Rajaram, and L. S. Schlesinger, "CD36-mediated uptake of surfactant lipids by human macrophages promotes intracellular growth of *Mycobacterium tuberculosis*," *The Journal of Immunology*, vol. 197, no. 12, pp. 4727–4735, 2016.

[161] F. L. Greene and L. H. Sobin, "The TNM system: Our language for cancer care," *Journal of Surgical Oncology*, vol. 80, no. 3, pp. 119–120, 2002.

[162] M. A. Hamburg and F. S. Collins, "The path to personalized medicine," *New England Journal of Medicine*, vol. 363, no. 4, pp. 301–304, 2010.

[163] I. Petrache, T. R. Medler, A. T. Richter, K. Kamocki, U. Chukwueke, L. Zhen, Y. Gu, J. Adamowicz, K. S. Schweitzer, W. C. Hubbard, E. V. Berdyshev, G. Lungarella, and R. M. Tuder, "Superoxide dismutase protects against apoptosis and alveolar enlargement induced by ceramide," *American Journal of Physiology. Lung Cellular and Molecular Physiology*, vol. 295, no. 1, pp. L44–L53, 2008.

[164] M. Bodas, T. Min, and N. Vij, "Lactosylceramide-accumulation in lipid-rafts mediate aberrant-autophagy, inflammation and apoptosis in cigarette smoke induced emphysema," *Apoptosis*, vol. 20, no. 5, pp. 725–739, 2015.

[165] S. Marwitz, S. Depner, D. Dvornikov, R. Merkle, M. Szczygiel, K. Muller-Decker, P. Lucarelli, M. Wasch, H. Mairbaurl, K. F. Rabe, C. Kugler, E. Vollmer, M. Reck, S. Scheufele, M. Kroger, O. Ammerpohl, R. Siebert, T. Goldmann, and U. Klingmuller, "Downregulation of the TGFbeta pseudoreceptor BAMBI in non-small cell lung cancer enhances TGFbeta signaling and invasion," *Cancer Res*, vol. 76, no. 13, pp. 3785–801, 2016.

[166] L. G. Dobbs, R. Gonzalez, M. A. Matthay, E. P. Carter, L. Allen, and A. S. Verkman, "Highly water-permeable type I alveolar epithelial cells confer high water permeability between the airspace and vasculature in rat lung," *Proceedings of the National Academy of Sciences*, vol. 95, no. 6, pp. 2991–2996, 1998.

[167] P. Mao, S. Wu, J. Li, W. Fu, W. He, X. Liu, A. S. Slutsky, H. Zhang, and Y. Li, "Human alveolar epithelial type II cells in primary culture," *Physiological Reports*, vol. 3, no. 2, 2015.

[168] A. L. Swan, D. J. Stekel, C. Hodgman, D. Allaway, M. H. Alqahtani, A. Mobasheri, and J. Bacardit, "A machine learning heuristic to identify biologically relevant and minimal biomarker panels from omics data," *BMC Genomics*, vol. 16 Suppl 1, p. S2, 2015.

[169] N. Lazzarini, P. Widera, S. Williamson, R. Heer, N. Krasnogor, and J. Bacardit, "Functional networks inference from rule-based machine learning models," *BioData Mining*, vol. 9, no. 1, p. 28, 2016.

[170] R. Herzog, D. Schwudke, K. Schuhmann, J. Sampaio, S. Bornstein, M. Schroeder, and A. Shevchenko, "A novel informatics concept for high-throughput shotgun lipidomics based on the molecular fragmentation query language," *Genome Biology*, vol. 12, no. 1, p. R8, 2011.

[171] M. C. Thomas, T. W. Mitchell, D. G. Harman, J. M. Deeley, J. R. Nealon, and S. J. Blanksby, "Ozone-induced dissociation: Elucidation of double bond position within mass-selected lipid ions," *Analytical Chemistry*, vol. 80, no. 1, pp. 303–311, 2008.

[172] X. Ma, L. Chong, R. Tian, R. Shi, T. Y. Hu, Z. Ouyang, and Y. Xia, "Identification and quantitation of lipid C=C location isomers: A shotgun lipidomics approach enabled by photochemical reaction," *Proceedings of the National Academy of Sciences*, vol. 113, no. 10, pp. 2573–2578, 2016.

[173] M. Kliman, J. C. May, and J. A. McLean, "Lipid analysis and lipidomics by structurally selective ion mobility-mass spectrometry," *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, vol. 1811, no. 11, pp. 935–945, 2011.

[174] J. A. Bowden, A. Heckert, C. Z. Ulmer, C. M. Jones, J. P. Koelmel, L. Abdullah, L. Ahonen, Y. Alnouti, A. Armando, J. M. Asara, T. Bamba, J. R. Barr, J. Bergquist, C. H. Borchers, J. Brandsma, S. B. Breitkopf, T. Cajka, A. Cazenave-Gassiot, A. Checa, M. A. Cinel, R. A. Colas, S. Cremers, E. A. Dennis, J. E. Evans, A. Fauland, O. Fiehn, M. S. Gardner, T. J. Garrett, K. H. Gotlinger, J. Han, Y. Huang, A. H. Neo, T. Hyotylainen, Y. Izumi, H. Jiang, H. Jiang, J. Jiang, M. Kachman, R. Kiyonami, K. Klavins, C. Klose, H. C. Kofeler, J. Kolmert, T. Koal, G. Koster, Z. Kuklenyik, I. J. Kurland, M. Leadley, K. Lin, K. R. Maddipati, D. McDougall, P. J. Meikle, N. A. Mellett, C. Monnin, M. A. Moseley, R. Nandakumar, M. Oresic, R. E. Patterson, D. Peake, J. S. Pierce, M. Post, A. D. Postle, R. Pugh, Y. Qui, O. Quehenberger, P. Ramrup, J. Rees, B. Rembiesa, D. Reynaud, M. R. Roth, S. Sales, K. Schuhmann, M. L. Schwartzman, C. N. Serhan, A. Shevchenko, S. E. Somerville, L. S. John-Williams, M. A. Surma, H. Takeda, R. Thakare, J. W. Thompson, F. Torta, A. Triebl, M. Trotzmuller, S. K. Ubhayasekera, D. Vuckovic, J. M. Weir, R. Welti, M. R. Wenk, C. E. Wheelock, L. Yao, M. Yuan, X. H. Zhao, and S. Zhou, "Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using standard reference material 1950 metabolites in frozen human plasma," *Journal of Lipid Research*, 2017.

[175] D. Schwudke, A. Shevchenko, N. Hoffmann, and R. Ahrends, "Lipidomics informatics for life-science," *Journal of Biotechnology*, 2017.

[176] C. M. Bernaards, J. W. R. Twisk, J. Snel, W. V. Mechelen, and H. C. G. Kemper, "Is calculating pack-years retrospectively a valid method to estimate life-time tobacco smoking? A comparison between prospectively calculated pack-years and retrospectively calculated pack-years," *Addiction*, vol. 96, pp. 1653–1661, 11 2001.

[177] L. F. Eggers and D. Schwudke, *Shotgun Lipidomics Approach for Clinical Samples*, pp. 163–174. New York, NY: Springer New York, 2018.

[178] X. Han, K. Yang, J. Yang, H. Cheng, and R. W. Gross, "Shotgun lipidomics of cardiolipin molecular species in lipid extracts of biological samples," *Journal of Lipid Research*, vol. 47, no. 4, pp. 864–879, 2006.

[179] G. Liebisch, J. A. Vizcaíno, H. Köfeler, M. Trötzmüller, W. J. Griffiths, G. Schmitz, F. Spener, and M. J. O. Wakelam, "Shorthand notation for lipid structures derived from mass spectrometry," *Journal of Lipid Research*, vol. 54, no. 6, pp. 1523–1530, 2013.

[180] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[181] G. Liebisch, M. Binder, R. Schifferer, T. Langmann, B. Schulz, and G. Schmitz, "High throughput quantification of cholesterol and cholesteryl ester by electrospray ionization tandem mass spectrometry (ESI-MS/MS)," *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, vol. 1761, no. 1, pp. 121–128, 2006.

[182] S. Lê, J. Josse, and F. Husson, "FactoMineR: An R package for multivariate analysis," *Journal of Statistical Software*, vol. 25, no. 1, pp. 1–18, 2008.

[183] T. Fawcett, "An introduction to ROC analysis," *Pattern Recogn. Lett.*, vol. 27, pp. 861–874, June 2006.

[184] M. de Hoon, S. Imoto, J. Nolan, and S. Miyano, "Open source clustering software," *Bioinformatics*, vol. 20, no. 9, pp. 1453–1454, 2004.

[185] A. J. Saldanha, "Java Treeview - Extensible visualization of microarray data," *Bioinformatics*, vol. 20, no. 17, pp. 3246–3248, 2004.

[186] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

[187] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller, "pROC: An open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, no. 1, p. 77, 2011.

[188] G. Sanchez, "plsdepot: Partial least squares (PLS) data analysis methods," 2012.

[189] G. Sanchez, "R package plsdepot PLS regression 2," *Documetation PLS depot*.

A Supplementary Information

This chapter contains:

- A list of patient related data: Tab. A.1, page 144.
- Scoring results for alveolar tissues: Tab. A.2, page 145.
- Scoring results for tumor tissues: Tab. A.3, page 147.
- A list of all identified lipid species: Tab. A.4, page 149.

Table A.4 contains all identified lipid species as used in the *follow-up study* with the *m/z* values and mass errors for each sub-cohort and occupation levels, mean values and SDs for the subsets of alveolar and tumor tissues. A respective list, specifically with the identified lipids from the *pilot study*, is online available with the publication [146] under the link `https://www.nature.com/articles/s41598-017-11339-1` (electronic supplement material, Supplement 3).

Table A.1: **Clinical data.** Age: age at surgery, NA: data not available, none: no COPD diagnosed.

Patient ID	Gender	Age	Cancer diagnosis	Pack-years	BMI	GOLD stage
2	male	71	SCC	65	35	1
4	male	66	SCC	100	25.14	1
6	male	70	SCC	50	27.77	1
11	male	48	ADC	NA	NA	NA
12	male	70	SCC	50	21.4	2
15	male	45	SCC	70	40.01	2
17	male	68	SCC	NA	NA	2
18	female	47	SCC	50	32	none
19	female	54	LCC	35	28	1
22	male	57	SCC	25	25	1
23	male	67	Carc	40	26.85	NA
24	male	55	ADC	40	25.5	none
26	male	69	Carc	NA	NA	NA
29	male	59	SCC	40	29	none
30	male	52	ADC	30	26.5	3
31	male	68	SCC	125	32.5	2
32	male	45	ADC	30	23.1	1
39	male	60	SCC	40	25.1	2
41	male	80	SCC	NA	25.65	NA
43	male	60	SCA	30	22.6	1
45	male	63	NET	50	22.05	2
49	female	75	ADC	18	29.22	NA
50	male	60	SCC	47	24.9	1
52	female	44	Carc	10	27	restriction
53	male	48	ADC	30	29	none
54	male	73	NET	NA	36.33	NA
55	NA	69	ADC	40	18.37	NA
58	male	51	SCC	60	18.83	1to2
61	male	70	ADC	60	18.25	2
62	male	56	SCC	45	29.05	NA
63	female	84	Carc	NA	NA	NA
64	female	51	ADC	35	26.96	1
65	female	58	NET	NA	NA	NA
66	male	62	ADC	45	28.4	1
67	female	52	ADC	50	29.68	2
71	female	57	ADC	30	20.66	1
72	male	60	ADC	40	33.5	2
76	female	75	LCC	22	23.51	1
85	female	74	ADC	10	22.55	2
87	female	40	Carc	NA	39.79	NA
91	female	76	ADC	NA	25.56	NA
93	female	67	LCC	40	23.67	2
94	male	65	SCC	120	32.11	NA
96	male	75	LCC	10	25.01	NA
97	male	54	SCC	NA	25.78	1
99	female	54	LCC	30	15.57	NA
100	male	56	SCC	80	25.7	NA
101	male	73	SCC	60	22.94	NA
102	female	78	SCA	NA	24.91	NA
103	female	73	ADC	never	27.58	NA
104	female	73	SCC	60	25.1	NA
105	female	67	ADC	never	18.9	NA
106	male	82	SCC	15	34.61	NA
107	female	75	ADC	NA	NA	NA
108	female	45	ADC	NA	22.27	NA
109	male	56	SCC	NA	24.15	2
111	male	60	ADC	NA	25.95	NA
113	female	66	ADC	50	18.42	3
114	male	75	ADC	30	29.26	2
118	female	71	ADC	NA	29.41	NA
119	male	45	ADC	35	30.19	NA
120	male	64	SCA	60	21.45	1
121	male	54	LCC	40	20.58	NA
122	male	51	Carc	never	27.47	NA
123	female	54	ADC	never	22.58	NA
125	male	51	SCC	33	26.23	2
126	male	71	SCC	NA	23.32	NA
127	female	55	ADC	55	20.69	2
128	female	66	ADC	30	25.86	NA
129	male	71	SCC	60	22.59	NA
130	female	62	ADC	15	26.35	2
133	male	71	SCC	NA	27.78	2
134	male	66	SCC	50	31.74	NA
136	female	56	ADC	NA	23.32	NA
138	female	77	ADC	NA	28.88	NA

continued on next page . . .

. . . continued from previous page

continued on next page . . .

. . . continued from previous page

Patient ID	Gender	Age	Cancer diagnosis	Pack-years	BMI	GOLD stage
140	female	73	ADC	60	NA	2
142	male	59	ADC	50	27.76	NA
143	male	75	SCC	20	28.73	NA
144	male	74	ADC	50	22.59	NA
145	male	73	ADC	NA	29.61	NA
146	male	52	SCC	50	26.17	NA
147	female	62	ADC	45	24.24	2
148	male	73	ADC	20	26.81	NA
149	male	75	ADC	NA	23.67	3
150	male	60	ADC	30	32.89	NA
151	male	64	ADC	NA	24.65	3
153	male	76	ADC	30	25.39	NA
154	male	62	ADC	NA	NA	NA
155	male	62	ADC	80	29.41	2
160	male	72	SCC	20	NA	NA
161	female	67	ADC	NA	NA	NA
164	male	67	SCC	80	25.35	2

Table A.2: **Scoring results from alveolar tissues.** NA: data not available.

Tissue ID	Emphysema	Inflammation	Alv. Macrophages	Fibrosis
ID2_A	5	1	2	2
ID4_A	3	1	1	1
ID6_A	5	2	3	2
ID11_A	4	1	2	1
ID12_A	4	1	1	2
ID15_A	3	1	2	1
ID17_A	5	2	1	2
ID18_A	4	1	2	2
ID19_A	1	1	2	3
ID22_A	3	2	2	1
ID23_A	NA	NA	NA	NA
ID24_A	2	3	3	2
ID26_A	NA	NA	NA	NA
ID29_A	3	2	2	2
ID39_A	2	1	1	1
ID41_A	NA	NA	NA	NA
ID43_A	4	2	1	2
ID45_A	NA	NA	NA	NA
ID49_A	NA	NA	NA	NA
ID50_A	3	2	2	2
ID52_A	3	1	1	2
ID53_A	3	1	1	1
ID54_A	NA	NA	NA	NA
ID55_A	NA	NA	NA	NA
ID58_A	NA	NA	NA	NA
ID61_A	4	1	2	1
ID62_A	NA	NA	NA	NA
ID63_A	NA	NA	NA	NA
ID64_A	3	1	1	1
ID65_A	NA	NA	NA	NA
ID66_A	3	1	2	1
ID67_A	3	1	1	2
ID76_A	NA	NA	NA	NA
ID85_A	NA	NA	NA	NA
ID87_A	NA	NA	NA	NA
ID91_A	5	2	1	2
ID93_A	NA	NA	NA	NA

continued on next page . . .

. . . continued from previous page

Tissue ID	Emphysema	Inflammation	Alv. Macrophages	Fibrosis
ID94_A	NA	NA	NA	NA
ID96_A	4	2	2	2
ID97_A	7	2	1	2
ID99_A	6	2	2	2
ID100_A	NA	NA	NA	NA
ID101_A	6	3	1	3
ID102_A	6	1	1	1
ID103_A	4	1	1	2
ID104_A	NA	NA	NA	NA
ID105_A	NA	NA	NA	NA
ID106_A	4	1	1	2
ID107_A	NA	NA	NA	NA
ID108_A	5	1	2	1
ID109_A	5	1	2	1
ID111_A	6	1	2	1
ID113_A	NA	NA	NA	NA
ID114_A	NA	NA	NA	NA
ID118_A	5	1	2	1
ID119_A	NA	NA	NA	NA
ID120_A	6	1	1	2
ID121_A	7	2	3	3
ID122_A	5	2	2	2
ID123_A	NA	NA	NA	NA
ID125_A	6	1	1	1
ID126_A	0	3	1	3
ID127_A	4	1	2	1
ID128_A	NA	NA	NA	NA
ID129_A	6	2	2	1
ID130_A	3	2	2	2
ID133_A	4	1	1	1
ID134_A	6	1	2	1
ID136_A	4	1	2	1
ID138_A	7	2	2	3
ID140_A	6	1	2	1
ID142_A	3	2	2	2
ID143_A	4	2	1	1
ID144_A	6	1	2	2
ID145_A	5	2	2	2
ID146_A	7	1	1	2
ID147_A	7	1	2	2
ID148_A	5	1	1	3
ID149_A	5	1	1	1
ID150_A	5	2	1	2
ID151_A	6	3	2	3
ID153_A	6	3	3	3
ID154_A	5	1	1	2
ID155_A	8	2	1	2
ID160_A	5	1	1	1
ID161_A	6	1	2	2
ID164_A	6	1	1	1

continued on next page . . .

Table A.3: **Scoring results from tumor tissues.** NA: data not available.

Tissue ID	Necrosis	Vital tumor	Stroma	Inflammation
ID2_T	0	60	40	NA
D2_T	0	60	40	2
ID4_T	20	30	50	3
ID6_T	20	40	40	2
ID12_T	0	50	50	2
ID18_T	30	40	30	2
ID19_T	70	20	10	1
ID22_T	0	10	90	3
ID23_T	NA	NA	NA	NA
ID24_T	100	0	0	0
ID26_T	NA	NA	NA	NA
ID29_T	10	80	10	2
ID30_T	20	30	50	3
ID31_T	30	30	40	2
ID32_T	0	20	80	1
ID39_T	10	50	40	3
ID41_T	NA	NA	NA	NA
ID43_T	10	80	10	1
ID45_T	NA	NA	NA	NA
ID49_T	NA	NA	NA	NA
ID50_T	30	40	30	3
ID53_T	10	20	70	2
ID54_T	NA	NA	NA	NA
ID55_T	NA	NA	NA	NA
ID58_T	NA	NA	NA	NA
ID61_T	10	20	70	2
ID62_T	NA	NA	NA	NA
ID63_T	NA	NA	NA	NA
ID64_T	0	10	90	3
ID65_T	NA	NA	NA	NA
ID66_T	10	40	50	2
ID67_T	10	30	60	1
ID71_T	10	60	30	2
ID72_T	0	50	50	1
ID76_T	NA	NA	NA	NA
ID85_T	NA	NA	NA	NA
ID87_T	NA	NA	NA	NA
ID91_T	0	20	80	2
ID93_T	NA	NA	NA	NA
ID94_T	NA	NA	NA	NA
ID96_T	30	40	30	3
ID97_T	0	20	80	3
ID99_T	30	60	10	2
ID100_T	0	50	50	1
ID101_T	0	30	70	2
ID102_T	0	100	0	1
ID103_T	0	50	50	2
ID104_T	NA	NA	NA	NA
ID105_T	NA	NA	NA	NA
ID106_T	70	20	10	2
ID107_T	NA	NA	NA	NA
ID108_T	10	60	30	3
ID109_T	10	80	10	2
ID111_T	40	30	30	3
ID113_T	NA	NA	NA	NA
ID114_T	NA	NA	NA	NA
ID118_T	0	40	60	2
ID119_T	NA	NA	NA	NA
ID120_T	10	80	10	2
ID121_T	20	40	40	3
ID122_T	0	90	10	1
ID123_T	0	40	60	3
ID125_T	no tumor - lymph node			
ID126_T	20	20	60	3
ID127_T	20	10	70	3
ID128_T	NA	NA	NA	NA
ID129_T	0	40	60	3
ID130_T	0	50	50	3
ID133_T	10	60	30	2
ID134_T	10	40	50	1
ID136_T	30	60	10	3
ID138_T	0	10	90	2
ID140_T	0	70	30	3
ID142_T	0	50	50	3
ID143_T	10	50	40	2
ID144_T	0	70	30	2
ID145_T	0	30	70	2

continued on next page . . .

. . . continued from previous page

continued on next page . . .

. . . continued from previous page

Tissue ID	Necrosis	Vital tumor	Stroma	Inflammation
ID146_T	10	50	40	2
ID147_T	0	30	70	3
ID148_T	0	70	30	1
ID149_T	0	80	20	1
ID150_T	20	40	50	2
ID151_T	0	30	70	3
ID153_T	10	60	30	3
ID154_T	0	50	50	2
ID155_T	0	40	60	1
ID160_T	20	40	40	2
ID161_T	NA	NA	NA	NA
ID164_T	NA	NA	NA	NA

Table A.4: **Identified lipid species.**

Species	Ion mode	*m/z* in sub-cohort		Error [ppm] in sub-cohort		Occupation [%]		Mean [pmol/mg tissue]		SD [pmol/mg tissue]	
		1	2	1	2	alveolar	tumor	alveolar	tumor	alveolar	tumor
CE 14:0	+	614.5867	614.5864	-0.5	-1	12.8	51.1	6.6	5.5	5.5	4.3
CE 15:0	+	628.6027	628.6023	0	-0.6	9.3	31.8	2.3	2.6	1.4	1.8
CE 16:0	+	642.6178	642.6176	-0.9	-1.2	86	96.6	28.4	55	37.4	56.8
CE 16:1	+	640.6026	640.6022	-0.2	-0.8	73.3	86.4	14.9	26.2	18.7	24.9
CE 17:1	+	654.6182	654.6178	-0.3	-0.8	11.6	42	4.9	4.7	3	3.6
CE 18:0	+	670.6509	670.6498	1.9	0.2	8.1	59.1	6.2	20.2	5.4	23.7
CE 18:1	+	668.6344	668.6336	0.6	-0.6	98.8	98.9	78.1	264.5	119.5	297.4
CE 18:2	+	666.6184	666.6179	0.1	-0.8	98.8	100	167.6	186.2	146.3	238.1
CE 18:4	+	662.5865	NA	-0.9	NA	7	0	2.9	NA	2.2	NA
CE 19:1	+	682.6493	682.649	-0.5	-1	1.2	36.4	6.3	4.3	NA	3.2
CE 19:2	+	680.6336	NA	-0.6	NA	1.2	8	1.3	2.8	NA	1.8
CE 20:1	+	696.6659	696.6651	0.8	-0.3	4.7	69.3	5.9	9.6	4.4	9.4
CE 20:2	+	694.6503	694.6494	0.9	-0.4	11.6	76.1	12.1	24.8	16.9	47.9
CE 20:3	+	692.6341	692.6337	0.1	-0.5	39.5	93.2	20.3	90.4	54.8	121.8
CE 20:4	+	690.6183	690.6179	-0.1	-0.7	95.3	97.7	26.8	188.6	41.5	310.6
CE 20:5	+	688.6022	688.6014	-0.7	-1.9	54.7	83	10	21.9	8.8	24.8
CE 22:1	+	724.6967	724.6962	0.1	-0.6	0	17	NA	3.1	NA	2.4
CE 22:2	+	722.6815	722.681	0.7	0	3.5	36.4	3.1	4.5	1.7	8.3
CE 22:3	+	NA	720.665	NA	-0.4	4.7	38.6	8.1	14.6	6.8	13.5
CE 22:4	+	718.6493	718.649	-0.5	-1	14	79.5	17.7	71.8	31	102.9
CE 22:5	+	716.6337	716.6333	-0.4	-1	12.8	81.8	10.8	38.8	19.8	52.4
CE 22:6	+	714.6184	714.6179	0.1	-0.7	26.7	80.7	7	50.6	13.8	110.8
CE 24:0	+	NA	754.7432	NA	-0.4	16.3	20.5	15.2	14.2	16.3	15.9
CE 24:1	+	752.728	752.7275	0.1	-0.5	15.1	63.6	13.3	15.8	16.1	15.7
CE 24:2	+	750.7128	750.7121	0.8	-0.2	1.2	29.5	6.2	4.7	NA	6.1
CE 24:4	+	746.681	746.6804	0	-0.7	4.7	61.4	8.7	11.5	6.7	13.8
CE 24:5	+	744.6658	744.6653	0.6	0	3.5	30.7	3.2	6	2.5	5.9
CE 26:1	+	NA	780.7596	NA	0.6	17.4	23.9	47.7	66.9	92.7	86.6
CE 26:2	+	NA	778.7431	NA	-0.6	17.4	19.3	100	54.1	96.3	43.5
CE 26:3	+	NA	776.7278	NA	-0.1	11.6	13.6	5.2	4	4.1	2.1
CE 26:4	+	774.7121	NA	-0.2	NA	0	6.8	NA	1.6	NA	1.1
CE 28:4	+	NA	802.7432	NA	-0.5	17.4	17	26.1	60.9	37.8	89.1
CE 30:4	+	NA	830.7751	NA	0.3	3.5	14.8	16.1	17	16.8	17.7
Cer 32:1;0	-	544.4498	544.4496	-0.8	-1	82.6	92	4.6	5.7	3	6.9
Cer 33:1;0	-	558.4652	558.4653	-1.2	-1.1	64	64.8	3.3	3.7	2	4
Cer 34:0;0	-	574.4967	574.497	-0.8	-0.3	81.4	92	4.5	7.6	2.8	7.3
Cer 34:1;0	-	572.4804	572.4809	-2	-1.1	100	100	121.4	225.9	91.9	438
Cer 34:1;1	-	588.4754	588.4759	-1.7	-0.9	2.3	17	1.1	1.6	0.6	0.8
Cer 34:2;0	-	570.4648	570.465	-1.9	-1.5	91.9	93.2	8	8.6	5.9	11.7
Cer 36:0;0	-	602.5282	602.5284	-0.5	-0.1	8.1	11.4	1.6	1.2	2.1	1.1
Cer 36:1;0	-	600.5124	600.5123	-0.7	-0.9	100	100	16.9	35.6	11.6	129.1
Cer 36:2;0	-	598.4957	598.4961	-2.4	-1.7	84.9	68.2	5	5.4	3.7	9
Cer 38:0;0	-	630.5598	630.5601	0.1	0.5	2.3	6.8	3.3	1.4	0.4	0.6
Cer 38:1;0	-	628.5435	628.5438	-1	-0.5	98.8	100	11	22.3	7.1	67.6
Cer 38:2;0	-	626.5275	626.5278	-1.5	-1.1	70.9	39.8	3.2	7.7	2.1	12.8
Cer 39:1;0	-	642.559	642.5595	-1.1	-0.5	48.8	52.3	2.3	7.6	1.4	18.2
Cer 40:1;0	-	656.5747	656.575	-1	-0.6	100	100	35.2	60.1	22.1	149.9
Cer 40:2;0	-	654.5588	654.5592	-1.5	-0.8	97.7	98.9	13.5	16.8	8.6	24.1
Cer 40:3;0	-	NA	652.544	NA	-0.2	17.4	14.8	2	3.5	0.9	2.5
Cer 41:1;0	-	670.5896	670.5902	-2.1	-1.3	97.7	100	12.6	28	7.7	57.7
Cer 41:2;0	-	668.575	668.5754	-0.6	0	93	96.6	6.5	10.5	4.3	17.2
Cer 42:1;0	-	684.6066	684.6064	-0.2	-0.4	100	100	57.5	82.4	40.2	105.3

continued on next page . . .

. . . continued from previous page

		m/z in sub-cohort		Error [ppm] in sub-cohort		Occupation [%]		Mean [pmol/mg tissue]		SD [pmol/mg tissue]	
Species	Ion mode	1	2	1	2	alveolar	tumor	alveolar	tumor	alveolar	tumor
Cer 42:2;0	-	682.59	682.5903	-1.5	-1.1	100	100	191.1	246.8	147.5	480.7
Cer 42:2;1	-	698.5845	698.5848	-2.2	-1.7	12.8	65.9	2.1	4.4	1.9	4.3
Cer 42:3;0	-	680.5756	680.576	0.3	0.9	100	100	30.5	33	19.3	46.5
Cer 43:1;0	-	698.6207	698.6213	-2.3	-1.5	19.8	20.5	2.4	2.4	1.4	1.6
Cer 43:2;0	-	696.6064	696.6071	-0.4	0.6	90.7	81.8	4.6	6.2	2.8	12.5
Cer 43:3;0	-	694.5916	NA	0.8	NA	2.3	4.5	3.6	3.1	0	3.4
Cer 44:2;0	-	710.6226	710.6223	0.4	0	73.3	77.3	3.9	6.6	2.7	9.5
Cer 44:3;0	-	708.6072	708.6076	0.7	1.2	29.1	29.5	2	2.9	1.2	3.6
CL 51:3	-	581.8553	NA	-0.1	NA	2.3	12.5	3.1	5.4	1.2	5.6
CL 68:5	-	698.9719	NA	-1.1	NA	0	10.2	NA	6.2	NA	6.3
CL 70:7	-	NA	710.9718	NA	-1.2	32.6	25	4.3	5.1	3.4	4
CL 72:5	-	727.0036	NA	-0.5	NA	4.7	18.2	3.9	6	1.4	2.9
CL 72:7	-	724.9881	724.9876	-0.4	-1	74.4	92	2.1	5.3	2.3	5.5
CL 72:8	-	723.9804	723.9807	-0.1	0.2	91.9	92	15.2	21.8	13.1	51
CL 74:10	-	735.9802	735.9811	-0.4	0.7	11.6	5.7	4.5	5	2.3	2.7
CL 74:8	-	737.9958	737.9963	-0.4	0.2	27.9	50	4	6.8	2.3	6.7
CL 78:18	-	NA	755.9484	NA	-1.1	34.9	6.8	5.2	2.9	5	2.1
CL 84:19	-	NA	796.9875	NA	-1	1.2	20.5	1.3	4.9	NA	2.7
DAG 32:0	+	586.5401	586.5401	-0.7	-0.7	7	19.3	1.8	2.3	1.8	2.2
DAG 32:1	+	584.5246	NA	-0.4	NA	2.3	13.6	1.5	1.8	1.5	2
DAG 34:1	+	612.5563	612.5557	0.2	-0.8	89.5	90.9	8.3	9.2	6.7	7.2
DAG 34:2	+	610.5401	610.5397	-0.6	-1.3	14	20.5	3.1	3.2	2.5	3.2
DAG 36:2	+	638.5718	638.5712	0.1	-0.9	84.9	86.4	8.7	8.3	8.3	6.4
DAG 36:3	+	636.5557	636.5552	-0.6	-1.5	18.6	21.6	7.7	4.7	7.3	3.1
DAG 36:4	+	634.5405	634.5399	0.1	-1	4.7	14.8	1.2	1.7	0.5	1.1
DAG 38:4	+	662.5712	662.5709	-1	-1.3	23.3	20.5	5.2	3.9	3.8	2.5
DAG 38:5	+	660.5555	660.5563	-0.9	0.3	1.2	14.8	1.3	2.1	NA	1.6
DAG 38:6	+	658.5403	NA	-0.4	NA	0	6.8	NA	0.9	NA	0.3
DAG 40:6	+	686.5716	NA	-0.2	NA	0	8	NA	1.4	NA	0.7
DAG 46:10	+	762.6027	NA	-0.5	NA	14	12.5	40.8	23	20.6	12.6
DAG 50:1	+	836.8077	NA	1.3	NA	0	11.4	NA	3.7	NA	6.5
DAG 50:2	+	834.79	NA	-1.1	NA	0	8	NA	4.1	NA	5.8
FC	QTOF (+)	446.1497	446.4656	-218.11	61.68	100	100	4579.2	6613	1865	5113
LCL 62:1	-	653.9668	NA	-0.9	NA	5.8	3.4	2.8	4.2	1.5	1.9
LPA 16:4	-	NA	401.174	NA	1.2	22.1	20.5	1.8	2.2	0.9	1.5
LPC 16:0	-	530.3014	530.3012	-0.9	-1.2	86	97.7	22.9	39.3	16.8	30
LPC 18:0	-	558.3323	558.3323	-1.5	-1.7	72.1	89.8	16	22.7	9	13.9
LPC 18:1	-	556.3167	556.3166	-1.5	-1.7	31.4	85.2	9	18.8	4.6	12.8
LPC 20:4	-	578.3019	578.3015	0	-0.7	9.3	34.1	11	12.7	6.7	9.6
LPE 18:0	-	480.309	480.3088	-1.2	-1.6	75.6	88.6	3.5	4.5	2.7	3.5
LPE 18:1	-	478.2934	478.2928	-1.1	-2.3	20.9	52.3	2.7	2.8	1.5	2.7
LPE 20:4	-	500.278	500.2779	-0.5	-0.8	34.9	48.9	5.2	6.5	5.2	11.6
LPG 16:4	-	475.2102	NA	-0.1	NA	5.8	8	2	6.1	0.5	5.2
LPG 18:0	-	511.3035	NA	-1.3	NA	5.8	1.1	4.1	0.9	3	NA
LPG 18:1	-	509.2877	509.2876	-1.5	-1.8	87.2	30.7	5.2	2	5.2	1.4
LPI 16:0	-	571.288	NA	-1.6	NA	1.2	5.7	3.2	1.5	NA	1.1
LPI 18:0	-	599.3192	599.3192	-1.7	-1.7	88.4	98.9	4.8	9.5	3.4	7.3
LPI 18:1	-	597.3041	597.3045	-0.8	-0.1	57	40.9	2.9	2.4	1.9	1.6
LPI 20:4	-	619.2885	619.2889	-0.6	0	70.9	80.7	2.8	5.5	1.8	5.9
LPS 18:0	-	524.2991	524.2992	-0.6	-0.5	18.6	26.1	3.6	2.9	2	2
MAG 22:2	+	428.3736	428.3739	0.3	1.2	0	6.8	NA	1.7	NA	2.1
PA 32:0	-	647.4647	647.4646	-1.6	-1.8	2.3	4.5	3.5	2.3	1.9	2.2
PA 32:5	-	637.3871	NA	-0.5	NA	7	2.3	2.9	1	1.3	0.7
PA 34:1	-	NA	673.4804	NA	-1.4	50	46.6	3.1	3	2	1.9

continued on next page . . .

. . . continued from previous page

Species	Ion mode	*m/z* in sub-cohort		Error [ppm] in sub-cohort		Occupation [%]		Mean [pmol/mg tissue]		SD [pmol/mg tissue]	
		1	2	1	2	alveolar	tumor	alveolar	tumor	alveolar	tumor
PA 34:2	-	671.4652	671.4653	-0.8	-0.7	23.3	15.9	2.4	2.1	1.9	1.8
PA 35:1	-	NA	687.497	NA	0	51.2	39.8	3.8	4.2	3.8	3.8
PA 35:2	-	NA	685.4803	NA	-1.6	23.3	18.2	3.7	2.8	3.1	1.9
PA 36:1	-	701.5114	701.5115	-1.8	-1.7	46.5	40.9	8.7	8.1	14.9	14.6
PA 36:2	-	699.497	699.4974	-0.1	0.5	43	42	3.6	3.9	3.8	7.3
PA 36:3	-	697.482	NA	0.9	NA	3.5	3.4	2.5	3.4	1.3	3.1
PA 36:6	-	691.4358	NA	1.9	NA	3.5	9.1	2.4	4.2	0.7	2.4
PA 37:2	-	NA	713.5131	NA	0.6	23.3	14.8	3.4	3.7	3	2
PA 37:4	-	NA	709.4805	NA	-1.3	14	8	3.1	3.6	2.4	2.4
PA 48:7	-	NA	857.6048	NA	-2.1	16.3	28.4	4.6	3.8	6.9	3.2
PC 28:0	-	712.4683	712.4685	-0.9	-0.7	94.2	52.3	6.5	4.2	6.1	4.7
PC 29:0	-	726.4844	726.4842	-0.2	-0.6	45.3	22.7	3.1	2.5	2.4	1.7
PC 30:0	-	740.4989	740.4995	-1.8	-1.1	100	100	289.6	98.9	258.8	94.8
PC 30:1	-	738.4842	738.4849	-0.5	0.5	96.5	75	11.3	8.8	11.8	9.1
PC 31:0	-	754.5152	754.5156	-0.9	-0.4	100	98.9	41.9	24.9	30.7	22.3
PC 31:1	-	752.4997	752.5002	-0.7	-0.1	81.4	64.8	4.2	4.9	3.6	4.7
PC 31:3	-	NA	748.4686	NA	-0.5	27.9	0	3.7	NA	3.2	NA
PC 32:0	-	768.5299	768.5303	-2.1	-1.7	100	100	2694	1075.5	2153.6	1482.3
PC 32:1	-	766.5147	766.5153	-1.6	-0.8	98.8	98.9	486.8	319.7	456	317.3
PC 32:2	-	764.4994	764.4997	-1.1	-0.7	96.5	92	16.6	17.3	18.8	17.6
PC 32:3	-	762.4838	762.4841	-1	-0.6	15.1	18.2	3.4	2.2	3.2	1.1
PC 33:0	-	782.5479	782.5475	1	0.3	100	95.5	26.1	12.9	17.2	13.4
PC 33:1	-	780.5326	780.5322	1.4	0.8	98.8	98.9	21.6	27.5	14.1	26.5
PC 33:2	-	778.517	778.5169	1.4	1.3	76.7	87.5	3.3	7.1	2.2	7.3
PC 33:4	-	774.4833	774.4833	-1.6	-1.7	40.7	55.7	2.4	3.6	1.5	3.8
PC 34:1	-	794.5466	794.5466	-0.7	-0.8	100	100	1623.8	2053.7	1298.9	1753
PC 34:2	-	792.5322	792.5323	0.8	1	100	100	528.5	598.8	295.5	593.9
PC 34:3	-	790.5178	790.5174	2.4	1.9	98.8	96.6	22.1	18.7	15.9	16.2
PC 34:4	-	788.5019	788.5014	2	1.4	23.3	23.9	4.6	5.7	4	4.5
PC 35:1	-	808.5622	808.5631	-0.9	0.3	100	98.9	10.1	19.2	7.1	18
PC 35:2	-	806.5462	806.5464	-1.3	-1	97.7	98.9	6.7	11.4	3.7	9.2
PC 35:4	-	802.5155	802.5159	-0.5	0	60.5	88.6	5.1	6.5	3.3	4.8
PC 35:5	-	800.4983	800.4985	-2.5	-2.2	72.1	86.4	3.6	5.8	2.2	5
PC 36:1	-	822.5778	822.5773	-0.9	-1.5	98.8	97.7	102.3	246.7	87.3	261.1
PC 36:2	-	820.5639	820.5634	1.3	0.6	100	100	322.8	615.4	183.9	540.5
PC 36:3	-	818.5482	818.5479	1.2	0.9	100	97.7	89.6	156.4	63.1	227.4
PC 36:4	-	816.5312	816.531	-0.4	-0.7	100	100	312.6	487.5	186.4	525.6
PC 36:5	-	814.515	814.5155	-1.2	-0.6	97.7	98.9	17	15.8	9.7	12.5
PC 37:4	-	830.5458	830.5465	-1.7	-0.8	45.3	77.3	3.8	5.9	1.9	4.6
PC 37:7	-	824.4985	824.4985	-2.1	-2.1	81.4	89.8	8.4	9.2	12.5	8.8
PC 38:2	-	848.5939	848.5941	-0.3	-0.1	14	73.9	3.1	17.5	1.5	20.5
PC 38:3	-	846.5804	846.5801	2.3	1.9	17.4	54.5	8.2	18.4	13.6	18.6
PC 38:4	-	844.5647	844.564	2.1	1.3	100	100	168.8	265.1	106	235.5
PC 38:5	-	842.5459	842.5463	-1.6	-1.1	100	98.9	66	114.6	32.7	83.1
PC 38:6	-	840.5337	840.5326	2.5	1.3	100	98.9	26.8	44	14.7	32
PC 39:9	-	848.4989	848.4989	-1.6	-1.6	81.4	79.5	5.3	3.6	3.8	2.7
PC 40:4	-	872.5941	872.5941	-0.1	-0.1	65.1	93.2	3.6	10.3	2.3	11.2
PC 40:6	-	868.5618	868.5618	-1.2	-1.2	97.7	100	8.6	21.8	4.8	15.4
PC 40:8	-	NA	864.533	NA	1.7	15.1	37.5	19.9	4.8	15.7	3.2
PC *O*- 30:0	-	NA	726.5208	NA	-0.2	17.4	15.9	1.4	2.8	1.3	2.9
PC *O*- 31:0	-	740.536	740.5365	-0.9	-0.3	100	100	29.2	36.9	15.9	42.7
PC *O*- 32:1	-	752.5364	752.5368	-0.3	0.2	95.3	80.7	15.4	24.9	9.2	23.3
PC *O*- 33:1	-	NA	766.5521	NA	-0.3	14	29.5	30.1	33	55.7	90.3
PC *O*- 34:1	-	780.569	780.5687	1.4	0.9	98.8	100	17.5	23.5	10.1	22.9
PC *O*- 34:2	-	778.5523	778.5531	0	1.1	61.6	92	4.9	22.4	3.5	23.1

continued on next page . . .

. . . continued from previous page

Species	Ion mode	*m/z* in sub-cohort		Error [ppm] in sub-cohort		Occupation [%]		Mean [pmol/mg tissue]		SD [pmol/mg tissue]	
		1	2	1	2	alveolar	tumor	alveolar	tumor	alveolar	tumor
PC *O*- 35:0	-	796.5988	796.5992	-0.6	-0.1	64	1.1	3.3	0.8	1.7	NA
PC *O*- 35:8	-	780.4725	NA	-2	NA	23.3	22.7	4.4	4	2.5	4
PC *O*- 36:1	-	808.5992	808.5994	-0.1	0.2	23.3	60.2	2.2	5.9	1.4	8.5
PC *O*- 36:10	-	790.4578	790.4577	-0.7	-0.9	69.8	70.5	7.6	10.1	5	7.5
PC *O*- 36:2	-	806.5825	806.5831	-1.4	-0.7	84.9	86.4	3.3	11.7	2.5	13.4
PC *O*- 36:4	-	802.5529	802.5533	0.8	1.2	100	94.3	25	20.8	21.4	19
PC *O*- 36:5	-	800.5366	800.5371	0	0.6	100	95.5	40.1	30.9	22.1	23.9
PC *O*- 37:10	-	804.4725	804.4721	-1.9	-2.4	51.2	62.5	3.2	3.5	2	2.5
PC *O*- 38:0	-	838.6471	838.6474	1	1.5	80.2	48.9	3.6	3.3	2.2	2.7
PC *O*- 38:1	-	NA	836.6297	NA	-1	3.5	19.3	5.2	5.2	0.7	4.6
PC *O*- 38:4	-	830.5829	830.583	-0.8	-0.7	75.6	72.7	6.7	12	4.7	13.1
PC *O*- 38:5	-	828.5687	828.5681	0.9	0.2	100	100	29.4	29.9	16.5	23.5
PC *O*- 38:6	-	826.5529	826.5536	0.7	1.6	96.5	92	10	12.6	5.1	8.7
PC *O*- 38:7	-	824.5378	824.5368	1.4	0.2	15.1	39.8	1.8	4.1	0.8	3.5
PC *O*- 39:1	-	NA	850.6484	NA	2.6	72.1	73.9	37.5	19.3	18	12.5
PC *O*- 40:1	-	NA	864.6625	NA	0.8	9.3	30.7	3	5.5	1.3	7
PC *O*- 40:4	-	NA	858.6159	NA	1.2	25.6	42	2.3	5	1.8	4.3
PC *O*- 40:6	-	NA	854.5856	NA	2.3	44.2	60.2	3.1	6	1.3	4.2
PC *O*- 40:7	-	NA	852.5681	NA	0.2	20.9	58	2.3	4.2	0.9	3
PC *O*- 40:8	-	NA	850.5537	NA	1.6	51.2	43.2	1.7	1.7	1.4	1.5
PC *O*- 42:2	-	NA	890.6781	NA	0.7	4.7	20.5	29.9	7.5	54.8	8.1
PC *O*- 42:3	-	NA	888.664	NA	2.5	3.5	21.6	5.6	4.4	7.2	4.3
PE 31:7	-	662.3812	662.3818	-2.2	-1.4	45.3	3.4	3.7	4.8	4.8	3.1
PE 32:0	-	690.5078	690.5076	-0.2	-0.5	95.3	65.9	3.5	2.8	1.6	1.7
PE 32:1	-	688.4915	688.4913	-1.1	-1.4	89.5	72.7	3.3	7.4	1.8	9.9
PE 33:3	-	NA	698.4752	NA	-2.1	20.9	6.8	7	2.5	10.9	1.3
PE 34:0	-	718.5398	718.5384	0.8	-1.2	89.5	51.1	8.7	6.8	4.7	5.7
PE 34:1	-	716.5238	716.5234	0.3	-0.2	100	100	50.2	80.4	20.7	83.5
PE 34:2	-	714.5074	714.5072	-0.7	-1	100	98.9	21.4	32.6	9.7	44.2
PE 35:1	-	730.5396	730.5398	0.5	0.8	61.6	85.2	2.4	5	1.1	4.2
PE 35:2	-	728.5236	728.5237	0.1	0.1	15.1	42	1.8	3.9	0.8	3.3
PE 36:1	-	744.554	744.5539	-1.1	-1.3	100	100	88.6	167.7	41.2	103.6
PE 36:2	-	742.5383	742.5385	-1.3	-1	100	98.9	93.8	221.4	43	241.6
PE 36:3	-	740.5222	740.5229	-1.8	-0.9	97.7	94.3	34.6	47.6	20.3	78.5
PE 36:4	-	738.5073	738.5079	-0.9	0	100	98.9	38.5	47.7	20.6	39.5
PE 36:5	-	NA	736.4908	NA	-2	69.8	60.2	3.7	5.6	2.1	7
PE 38:1	-	772.5858	772.5855	-0.5	-0.9	72.1	85.2	2.7	3.7	1.1	2.5
PE 38:2	-	770.5693	NA	-1.6	NA	16.3	22.7	3	10.1	1.5	10.3
PE 38:3	-	768.5559	768.5562	1.3	1.7	95.3	98.9	16.9	40.6	7.4	47.3
PE 38:4	-	766.5376	766.5382	-2.1	-1.3	98.8	97.7	248.5	431.4	149	358.7
PE 38:5	-	764.5227	764.5228	-1.2	-1	100	98.9	60.2	95.7	30.2	83.1
PE 38:6	-	762.5074	762.5073	-0.7	-0.8	96.5	98.9	13.5	21	6.6	15.7
PE 40:2	-	798.6002	NA	-2	NA	2.3	11.4	4	3.5	0.7	2.4
PE 40:3	-	796.5876	NA	1.7	NA	0	8	NA	5.1	NA	3.5
PE 40:5	-	792.5567	792.5568	2.3	2.4	72.1	95.5	6.2	30.8	4.3	30.9
PE 40:6	-	790.5391	790.539	-0.2	-0.3	75.6	97.7	14.3	43.3	13.9	30
PE *O*- 31:7	-	662.3812	662.3818	-2.2	-1.4	45.3	3.4	3.7	4.8	4.8	3.1
PE *O*- 32:0	-	690.5078	690.5076	-0.2	-0.5	95.3	65.9	3.5	2.8	1.6	1.7
PE *O*- 32:1	-	688.4915	688.4913	-1.1	-1.4	89.5	72.7	3.3	7.4	1.8	9.9
PE *O*- 33:3	-	NA	698.4752	NA	-2.1	20.9	6.8	7	2.5	10.9	1.3
PE *O*- 34:0	-	718.5398	718.5384	0.8	-1.2	89.5	51.1	8.7	6.8	4.7	5.7
PE *O*- 34:1	-	716.5238	716.5234	0.3	-0.2	100	100	50.2	80.4	20.7	83.5
PE *O*- 34:2	-	714.5074	714.5072	-0.7	-1	100	98.9	21.4	32.6	9.7	44.2
PE *O*- 35:1	-	730.5396	730.5398	0.5	0.8	61.6	85.2	2.4	5	1.1	4.2

continued on next page . . .

. . . continued from previous page

Species	Ion mode	*m/z* in sub-cohort		Error [ppm] in sub-cohort		Occupation [%]		Mean [pmol/mg tissue]		SD [pmol/mg tissue]	
		1	2	1	2	alveolar	tumor	alveolar	tumor	alveolar	tumor
PE *O*- 35:2	-	728.5236	728.5237	0.1	0.1	15.1	42	1.8	3.9	0.8	3.3
PE *O*- 36:1	-	744.554	744.5539	-1.1	-1.3	100	100	88.6	167.7	41.2	103.6
PE *O*- 36:2	-	742.5383	742.5385	-1.3	-1	100	98.9	93.8	221.4	43	241.6
PE *O*- 36:3	-	740.5222	740.5229	-1.8	-0.9	97.7	94.3	34.6	47.6	20.3	78.5
PE *O*- 36:4	-	738.5073	738.5079	-0.9	0	100	98.9	38.5	47.7	20.6	39.5
PE *O*- 36:5	-	NA	736.4908	NA	-2	69.8	60.2	3.7	5.6	2.1	7
PE *O*- 38:1	-	772.5858	772.5855	-0.5	-0.9	72.1	85.2	2.7	3.7	1.1	2.5
PE *O*- 38:2	-	770.5693	NA	-1.6	NA	16.3	22.7	3	10.1	1.5	10.3
PE *O*- 38:3	-	768.5559	768.5562	1.3	1.7	95.3	98.9	16.9	40.6	7.4	47.3
PE *O*- 38:4	-	766.5376	766.5382	-2.1	-1.3	98.8	97.7	248.5	431.4	149	358.7
PE *O*- 38:5	-	764.5227	764.5228	-1.2	-1	100	98.9	60.2	95.7	30.2	83.1
PE *O*- 38:6	-	762.5074	762.5073	-0.7	-0.8	96.5	98.9	13.5	21	6.6	15.7
PE *O*- 40:2	-	798.6002	NA	-2	NA	2.3	11.4	4	3.5	0.7	2.4
PE *O*- 40:3	-	796.5876	NA	1.7	NA	0	8	NA	5.1	NA	3.5
PE *O*- 40:5	-	792.5567	792.5568	2.3	2.4	72.1	95.5	6.2	30.8	4.3	30.9
PE *O*- 40:6	-	790.5391	790.539	-0.2	-0.3	75.6	97.7	14.3	43.3	13.9	30
PE *O*- 32:1	-	674.5119	674.5119	-1.6	-1.6	64	56.8	2.3	3.8	1	3.1
PE *O*- 32:2	-	672.4976	672.4977	0.4	0.6	7	56.8	1.2	5.3	0.4	5.1
PE *O*- 34:1	-	702.5453	702.5455	1.4	1.7	69.8	79.5	3.4	5.6	2.5	4.4
PE *O*- 34:2	-	700.528	700.5285	-0.9	-0.3	100	100	53.8	90.3	27.8	73.6
PE *O*- 34:3	-	698.5113	698.5121	-2.4	-1.3	96.5	98.9	6.7	21.6	3.6	23.1
PE *O*- 35:2	-	714.5437	714.5435	-0.9	-1.2	68.6	80.7	3.2	5	1.7	4
PE *O*- 35:3	-	712.5272	712.5276	-2	-1.5	0	33	NA	2.7	NA	1.6
PE *O*- 36:1	-	730.5762	730.5761	0.8	0.6	17.4	37.5	2.3	4.4	0.7	3.2
PE *O*- 36:2	-	728.5604	728.5602	0.6	0.3	100	100	19.5	36.2	10.1	35.6
PE *O*- 36:3	-	726.5443	726.5441	-0.1	-0.3	98.8	100	18.6	42.5	9.5	39.3
PE *O*- 36:4	-	724.5279	724.5282	-1	-0.6	100	100	13.7	31	7.2	23.6
PE *O*- 36:5	-	722.5123	722.5123	-1	-1	100	100	339.4	328.9	228.4	240
PE *O*- 38:2	-	756.5903	756.5909	-1.2	-0.5	54.7	64.8	2.2	4.7	1.1	4.6
PE *O*- 38:3	-	754.5748	754.5757	-1.1	0.2	50	70.5	2.8	6.8	1.3	5.6
PE *O*- 38:5	-	750.5435	750.5435	-1.1	-1.1	98.8	100	468.5	357.2	351.7	288.2
PE *O*- 38:6	-	748.5292	748.5285	0.8	-0.2	76.7	96.6	182.2	160.1	214.6	125.6
PE *O*- 38:7	-	746.5137	746.5127	0.9	-0.4	98.8	98.9	43.9	64.6	27.6	46.1
PE *O*- 40:2	-	NA	784.6231	NA	0.7	2.3	19.3	3.5	3.4	3.1	2.6
PE *O*- 40:3	-	782.6077	782.6073	1.1	0.5	7	42	2.1	4.4	0.7	3.5
PE *O*- 40:4	-	780.5928	780.5922	1.9	1.2	10.5	45.5	0.2	2.9	0.2	4.6
PE *O*- 40:5	-	778.5772	778.577	2.1	1.8	100	98.9	28.2	39.1	14.6	44.7
PE *O*- 42:3	-	810.6391	810.6388	1.1	0.7	0	9.1	NA	4.1	NA	2.1
PE *O*- 42:5	-	806.606	806.6065	-1.1	-0.5	15.1	45.5	1.9	4.4	0.8	5.5
PE *O*- 42:6	-	NA	804.5925	NA	1.5	26.7	54.5	2.8	7.1	2	6.6
PE *O*- 43:10	-	NA	810.5453	NA	1.2	32.6	38.6	23.1	13.4	59.3	30.4
PG 30:0	-	693.4717	693.4715	0.7	0.5	90.7	29.5	6.6	4	6.5	3.2
PG 30:1	-	691.4564	NA	1.2	NA	8.1	3.4	2.9	1.2	1.3	0.8
PG 31:0	-	707.4854	707.4867	-2.1	-0.2	45.3	6.8	2.9	1.6	2.3	0.6
PG 32:0	-	721.5017	721.5019	-1.2	-0.8	100	78.4	84.7	20.4	86.3	30
PG 32:1	-	719.4857	719.4861	-1.7	-1.1	100	62.5	43.9	15	42.2	17.4
PG 32:2	-	NA	717.4711	NA	-0.2	31.4	5.7	3.3	2.5	2.9	1.1
PG 33:1	-	733.5016	733.5016	-1.2	-1.2	89.5	25	5.3	2.5	4.8	1.5
PG 34:0	-	749.533	749.5333	-1.1	-0.7	89.5	23.9	57.3	18.1	72.8	27.5
PG 34:1	-	747.5175	747.5182	-0.9	0	100	100	1124.8	167.6	1028.5	254.5
PG 34:2	-	745.5032	745.5037	0.9	1.6	100	97.7	142.3	40.3	166.2	74.4
PG 34:3	-	743.4876	743.4882	1	1.8	53.5	20.5	11.4	9.3	9.8	10
PG 35:1	-	761.5343	761.5345	0.7	0.9	96.5	75	17.4	3.7	15.5	3.3
PG 35:2	-	759.5185	759.5191	0.5	1.2	93	39.8	7.9	4.2	8.3	3.7

continued on next page . . .

. . . continued from previous page

		m/z in sub-cohort		Error [ppm] in sub-cohort		Occupation [%]		Mean [pmol/mg tissue]		SD [pmol/mg tissue]	
Species	Ion mode	1	2	1	2	alveolar	tumor	alveolar	tumor	alveolar	tumor
PG 36:0	-	777.5644	777.5635	-0.9	-2.1	95.3	40.9	7.2	2.6	8	4.2
PG 36:1	-	775.5486	775.5484	-1.1	-1.3	100	98.9	752.9	77.1	808.1	116.2
PG 36:2	-	773.5327	773.5333	-1.5	-0.7	100	100	1487.9	219.7	1582.4	344.7
PG 36:3	-	771.5185	771.5189	0.5	0.9	91.9	84.1	273.4	70.9	450.7	104.3
PG 36:4	-	769.5041	NA	2.1	NA	24.4	19.3	172.1	65.7	230.5	84.1
PG 38:1	-	803.5795	803.5797	-1.6	-1.3	75.6	40.9	5.4	2.1	5.5	1.7
PG 38:2	-	801.5653	801.5649	0.3	-0.3	95.3	69.3	6.2	3.9	5.7	3.7
PG 38:3	-	799.5508	799.55	1.7	0.7	17.4	20.5	2.2	7	1.8	6.6
PG 38:4	-	797.5356	797.5354	2.2	2	68.6	33	74.9	43.2	120.8	48.1
PG 38:5	-	795.5191	795.5194	1.2	1.6	86	42	136.9	58.7	271.8	61.2
PG 38:6	-	793.5026	793.5027	0.1	0.3	97.7	80.7	54.1	18.7	125.4	24.2
PG 38:7	-	791.4856	791.4855	-1.6	-1.7	48.8	19.3	5.7	3.7	9.6	4.9
PG 40:3	-	NA	827.5792	NA	-1.9	8.1	21.6	2.7	2.6	1.8	1.1
PG 40:4	-	825.5652	NA	0.1	NA	7	4.5	6.7	8.7	5.3	5.2
PG 40:6	-	821.5352	821.5351	1.7	1.6	97.7	68.2	53.1	26	68.9	23.9
PG 40:7	-	819.5176	819.5174	-0.7	-0.9	91.9	84.1	109.1	46.6	200.1	56.1
PG 41:1	-	845.6295	NA	2.1	NA	24.4	25	41.9	30.6	23.4	28.5
PG 42:0	-	NA	861.661	NA	2.3	69.8	51.1	7.9	4	5.4	2.6
PG 43:0	-	NA	875.6767	NA	2.3	66.3	42	5.7	4.5	3.9	2.7
PI 32:0	-	NA	809.519	NA	0.6	69.8	58	6.9	11.9	4.3	11.9
PI 32:1	-	807.5048	807.504	2.4	1.4	80.2	83	6.7	13.9	4.9	30.3
PI 34:0	-	837.5515	NA	1.9	NA	24.4	23.9	31.3	40.3	21.1	43.1
PI 34:1	-	835.5353	835.5351	1.3	1.1	100	100	212.2	215.8	193.2	390.5
PI 34:2	-	833.5175	833.518	-1.3	-0.6	98.8	98.9	27.9	55.7	17.9	101.7
PI 34:8	-	NA	821.4248	NA	0.2	0	31.8	NA	3.7	NA	2.5
PI 35:1	-	849.5504	849.5507	0.6	1.1	81.4	71.6	5.2	6.9	3.3	7.5
PI 35:2	-	NA	847.5359	NA	2	5.8	14.8	3.7	4.2	2.8	2.8
PI 36:1	-	863.5668	863.5672	1.5	2	100	100	368.6	237.6	390.4	281.6
PI 36:2	-	861.5492	861.5491	-0.8	-0.9	100	100	359.3	285.7	345.2	325.8
PI 36:3	-	859.5355	859.5342	1.5	0	100	97.7	25.2	43.6	16.2	45.5
PI 36:4	-	857.5201	857.5195	1.8	1.1	100	98.9	49.5	139.5	26.8	152
PI 36:5	-	855.5045	855.5047	1.9	2.1	94.2	100	8	11	4.2	6.9
PI 37:2	-	875.5647	NA	-0.9	NA	1.2	8	1.9	4.5	NA	3.7
PI 37:3	-	NA	873.549	NA	-1	4.7	36.4	1	3.5	1	4.2
PI 37:4	-	871.5334	871.5335	-0.9	-0.8	82.6	96.6	4.4	10.4	2.3	9.3
PI 38:4	-	885.5488	885.5485	-1.2	-1.5	98.8	100	1451.5	2969.2	1089.7	2537.5
PI 38:5	-	883.5353	883.5343	1.2	0.1	100	98.9	80.1	154.3	47.8	168.5
PI 38:6	-	881.5195	881.5198	1	1.4	100	98.9	17.6	18.8	9.6	15
PI 39:4	-	899.5677	899.5668	2.5	1.4	36	61.4	2.1	5	0.9	3.8
PI 40:4	-	913.5807	913.5807	-0.5	-0.5	100	98.9	11.4	47.7	6.5	53.5
PI 40:6	-	909.5488	909.5488	-1.2	-1.2	100	98.9	18.8	32.8	11	22.5
PI 40:7	-	907.5349	907.5365	0.7	2.6	74.4	76.1	3.4	3.5	1.9	2.4
PI 40:8	-	905.517	905.519	-1.7	0.5	58.1	64.8	3.1	3.2	1.5	1.8
PS 32:0	-	734.4973	734.4978	-0.7	0	51.2	19.3	1.9	3.2	1.6	4.1
PS 32:1	-	732.4826	732.483	0.7	1.3	52.3	21.6	2.6	5.4	1.5	7.5
PS 33:0	-	748.5121	NA	-1.7	NA	7	3.4	27.7	10.7	14.1	1.6
PS 33:8	-	NA	732.3898	NA	2.2	32.6	0	3.2	NA	3.7	NA
PS 34:0	-	762.5283	NA	-1	NA	4.7	4.5	2.6	3	2.6	1.3
PS 34:1	-	760.5126	760.5125	-1	-1.2	100	97.7	54.9	39.7	31.8	64.7
PS 34:2	-	758.4974	758.4975	-0.5	-0.3	89.5	40.9	7.7	6.9	4.6	7.6
PS 36:1	-	788.5434	788.5435	-1.6	-1.6	100	100	2126.6	1184.7	1970.8	799.9
PS 36:2	-	786.528	786.5286	-1.4	-0.6	100	98.9	164.2	129.7	123.8	103.2
PS 36:3	-	784.5124	784.5129	-1.3	-0.7	97.7	80.7	10.5	6.8	6.1	6
PS 36:4	-	782.4982	782.4975	0.6	-0.3	95.3	43.2	9.3	5.1	7.2	5.5
PS 38:1	-	816.5758	816.5759	-0.3	-0.2	100	98.9	27.2	18	17.1	16

continued on next page . . .

. . . continued from previous page

Species	Ion mode	*m/z* in sub-cohort		Error [ppm] in sub-cohort		Occupation [%]		Mean [pmol/mg tissue]		SD [pmol/mg tissue]	
		1	2	1	2	alveolar	tumor	alveolar	tumor	alveolar	tumor
PS 38:2	-	814.5599	814.5605	-0.5	0.2	100	97.7	31.2	28.6	21.9	30.3
PS 38:3	-	NA	812.5464	NA	2.1	75.6	75	78.3	109.9	46.7	165.6
PS 38:4	-	810.53	810.5298	1.2	1	100	100	215.6	109.7	181.9	87.3
PS 38:5	-	808.5135	808.5142	0.1	0.9	81.4	35.2	12.5	6.9	7.9	3.6
PS 40:1	-	NA	844.6092	NA	2.3	75.6	71.6	18.4	11.6	11.4	9.1
PS 40:2	-	842.5903	842.5908	-1.6	-1	100	98.9	34.1	21.4	24.7	22.8
PS 40:4	-	838.5617	838.5617	1.6	1.6	100	98.9	51.1	102.2	39	224.3
PS 40:6	-	834.5295	834.5287	0.6	-0.5	100	100	49.5	75.5	32.3	58.6
PS 40:7	-	832.5139	NA	0.5	NA	8.1	5.7	3.7	3.4	1.2	3
PS 41:0	-	860.6398	860.6377	1.4	-1.1	7	54.5	1.7	7.7	0.8	8.5
PS 42:1	-	872.6381	872.6384	-0.6	-0.3	97.7	89.8	9.3	6.4	6.7	5.2
PS 42:2	-	NA	870.6252	NA	2.5	73.3	68.2	8.2	7.6	4.9	7.8
PS 42:3	-	868.6066	868.6063	-0.8	-1.1	59.3	56.8	3.1	3.7	1.9	2.7
PS 42:4	-	NA	866.5934	NA	2	44.2	46.6	3.5	5.7	2.9	6.6
PS 43:4	-	880.6092	NA	2.1	NA	3.5	8	2	2.6	0.5	1.5
SM 32:1;0	+	675.5433	675.543	-0.4	-0.8	90.7	87.5	21.5	32.4	11.5	25.6
SM 34:0;0	+	705.5919	705.5906	2	0.1	60.5	61.4	25.9	38.8	9.4	23.6
SM 34:1;0	+	703.5751	703.5745	0.3	-0.5	100	100	562.8	639	209.8	420.2
SM 34:2;0	+	701.559	701.5585	-0.2	-1.1	91.9	79.5	19.7	27	8.2	25
SM 36:1;0	+	731.6059	731.6054	-0.4	-1.1	98.8	89.8	74.8	43.6	30.1	34
SM 36:2;0	+	729.5902	729.5899	-0.4	-0.9	7	10.2	14.2	11	4.5	9
SM 38:1;0	+	759.6383	759.6376	1.1	0.2	98.8	76.1	49.8	24.9	21	17.5
SM 38:2;0	+	757.6216	757.6217	-0.2	-0.1	29.1	10.2	9.4	16.2	3.8	11.4
SM 40:1;0	+	787.6694	787.669	0.8	0.4	98.8	96.6	135	56.2	58.7	39.3
SM 40:2;0	+	785.6527	785.6527	-0.5	-0.6	98.8	78.4	40.5	29.3	19.4	28.5
SM 41:1;0	+	801.6839	801.6841	-0.6	-0.4	98.8	80.7	37.9	31.5	17.3	27.5
SM 41:2;0	+	799.6696	799.6699	1.1	1.4	20.9	15.9	13.8	12.8	9.5	10.6
SM 42:1;0	+	815.7004	815.6999	0.4	-0.2	100	100	203.1	91.5	99.5	58.9
SM 42:2;0	+	813.6849	813.6849	0.6	0.6	100	100	392.5	223.3	167.4	147.3
SM 42:3;0	+	811.6686	811.6685	-0.2	-0.3	98.8	92	67.3	46.1	29.5	41.7
SM 42:4;0	+	809.6512	NA	-2.3	NA	8.1	5.7	14	9.6	11.5	4.5
SM 43:2;0	+	827.7015	NA	1.8	NA	16.3	3.4	9.6	8.3	4	4.4
SM 44:4;0	+	837.6827	837.6825	-2	-2.2	47.7	23.9	25.7	13.6	16.2	7.2
TAG 37:3	+	664.5506	664.5504	-0.7	-1	45.3	67	25.9	27.1	40	31.8
TAG 37:4	+	662.5348	662.5346	-1	-1.2	32.6	50	14.2	13.7	20.6	16.7
TAG 39:4	+	690.5668	690.5663	0.1	-0.6	58.1	89.8	45.8	86	117.5	166.9
TAG 39:5	+	688.5508	688.5505	-0.4	-0.8	81.4	84.1	56	46.7	62.6	61.7
TAG 41:5	+	NA	716.582	NA	-0.5	3.5	27.3	11.2	12.4	4.1	20.7
TAG 41:6	+	714.5666	714.5662	-0.2	-0.7	15.1	63.6	18.3	21.7	35.1	33.8
TAG 41:7	+	712.5514	712.5508	0.4	-0.3	45.3	77.3	24.9	58.7	39.4	91.3
TAG 41:8	+	NA	710.5353	NA	-0.2	18.6	27.3	5.9	14.6	7	12.5
TAG 42:1	+	738.6611	NA	0.6	NA	2.3	9.1	14.1	9	18.4	19.4
TAG 43:8	+	738.5673	738.5667	0.8	0	3.5	37.5	16.6	19.9	13.9	23
TAG 43:9	+	NA	736.5512	NA	0.2	1.2	26.1	30.3	30.3	NA	44.5
TAG 44:1	+	766.6924	766.6917	0.7	-0.3	25.6	43.2	28.6	18.8	87.1	59.8
TAG 44:16	+	NA	736.4574	NA	0.3	24.4	12.5	14.1	8.4	9.7	5.3
TAG 44:2	+	764.6753	764.6752	-1.2	-1.4	12.8	22.7	25.5	14.5	56.5	38.7
TAG 45:4	+	NA	774.6603	NA	-0.4	4.7	25	11.5	13.7	7.4	12.6
TAG 46:0	+	796.7392	796.7391	0.5	0.3	34.9	52.3	20	16.2	75.9	36.2
TAG 46:1	+	794.7231	794.7227	-0.1	-0.7	54.7	71.6	46.1	38.5	204.4	134.4
TAG 46:2	+	792.7076	792.7075	0.1	-0.1	23.3	51.1	50.1	23.5	154.9	73.9
TAG 46:3	+	790.6912	790.691	-0.9	-1.2	9.3	22.7	25.9	11.3	52.7	26.4
TAG 48:1	+	822.7557	822.7553	1.5	1	97.7	94.3	88.9	74.9	534.4	245.9
TAG 49:1	+	836.7712	836.7708	1.3	0.8	31.4	69.3	26.1	11.3	99.5	19.2
TAG 49:2	+	834.7536	834.7539	-1.1	-0.8	19.8	37.5	26.8	11.4	82.8	19.7

continued on next page . . .

. . . continued from previous page

Species	Ion mode	*m/z* in sub-cohort		Error [ppm] in sub-cohort		Occupation [%]		Mean [pmol/mg tissue]		SD [pmol/mg tissue]	
		1	2	1	2	alveolar	tumor	alveolar	tumor	alveolar	tumor
TAG 50:1	+	850.785	850.7855	-0.9	-0.3	100	98.9	182.3	175.8	1127.8	448.4
TAG 50:2	+	848.7711	848.771	1.1	1	100	100	217.4	144.7	1409.2	469.8
TAG 50:3	+	846.7546	846.7547	0.1	0.3	87.2	89.8	78.6	52.2	463.1	171.6
TAG 50:4	+	844.7379	844.7384	-1.1	-0.5	33.7	65.9	33.5	27.5	130.3	94
TAG 50:5	+	842.7244	842.7237	1.4	0.5	18.6	33	8.1	13.6	19.5	36
TAG 51:1	+	864.8024	864.8023	1.1	1	31.4	70.5	14	10	55.9	13.6
TAG 51:2	+	862.7846	862.7855	-1.5	-0.4	48.8	77.3	32.1	17.3	147.9	34.6
TAG 51:3	+	860.7703	860.771	0.2	1	16.3	45.5	35.4	9.6	101.2	17.6
TAG 51:4	+	NA	858.7558	NA	1.5	3.5	18.2	52	9	87.1	9.1
TAG 52:1	+	878.8161	NA	-1.1	NA	7	21.6	3	29	2.1	35.5
TAG 52:2	+	876.8013	876.8018	-0.2	0.4	100	100	937.1	318.9	7088.1	913.9
TAG 52:3	+	874.7853	874.7859	-0.6	0.1	100	100	604.3	213.5	4385.2	667.6
TAG 52:4	+	872.7717	872.7716	1.7	1.6	97.7	96.6	107.4	88.9	636.4	261.9
TAG 52:5	+	870.7546	870.7553	0.1	0.9	52.3	78.4	29.9	35.5	116.9	115.8
TAG 52:6	+	868.7378	868.7379	-1.2	-1.1	11.6	51.1	17.9	13.7	40.1	37.7
TAG 53:1	+	892.834	892.8335	1.3	0.8	5.8	39.8	1.2	4.8	0.7	6.3
TAG 53:2	+	890.817	890.8176	-0.1	0.5	37.2	69.3	26.2	14.3	114.3	23.9
TAG 53:3	+	888.8009	888.8017	-0.6	0.3	27.9	56.8	38	15.6	146.9	29.2
TAG 53:4	+	886.787	886.7874	1.4	1.8	9.3	46.6	32.2	11	74.4	18.2
TAG 53:5	+	884.7717	NA	1.8	NA	2.3	5.7	3.5	3.3	2.9	2
TAG 54:2	+	904.8322	904.8315	-0.7	-1.4	97.7	97.7	57.9	88.8	363	222.3
TAG 54:3	+	902.8176	902.8183	0.5	1.3	100	100	304.3	174.4	2188.5	490.2
TAG 54:4	+	900.8036	900.8034	2.4	2.2	98.8	100	154.1	180.1	1023.3	743.9
TAG 54:5	+	898.7862	898.7864	0.4	0.7	95.3	96.6	60.7	169.9	315.4	694.1
TAG 54:6	+	896.769	896.7701	-1.3	-0.1	19.8	25	9	34.8	13.4	47.6
TAG 54:7	+	894.7548	894.7556	0.4	1.2	5.8	17	3.3	7.8	3.7	8.7
TAG 55:1	+	920.8636	NA	-0.5	NA	0	9.1	NA	1.7	NA	1.1
TAG 55:2	+	918.8495	918.8498	1.2	1.5	9.3	33	13.8	11.2	28.8	14.4
TAG 55:3	+	916.8345	916.8348	1.9	2.3	9.3	37.5	21	12.4	47.1	20.1
TAG 55:4	+	914.8167	914.8179	-0.4	0.8	3.5	45.5	2	12.2	0.8	21.4
TAG 55:5	+	912.8002	912.8014	-1.3	-0.1	0	39.8	NA	13.3	NA	23.6
TAG 55:6	+	910.7869	NA	1.2	NA	1.2	9.1	1.5	3.6	NA	3
TAG 56:0	+	936.8948	936.8945	-0.6	-0.9	7	27.3	6.6	4.1	8	4.3
TAG 56:2	+	932.8641	932.8648	0	0.8	36	70.5	13.3	20.6	51.9	37.3
TAG 56:3	+	930.8474	930.8486	-1.1	0.2	46.5	79.5	20.9	24.5	96.1	57.4
TAG 56:4	+	928.8331	928.8344	0.3	1.8	47.7	88.6	17.2	77.7	66.5	217.6
TAG 56:6	+	924.802	924.8031	0.5	1.8	73.3	93.2	14	112.7	44	328.6
TAG 56:7	+	922.785	922.786	-0.9	0.3	64	92	14.7	70.7	48.8	214.4
TAG 56:8	+	920.7698	920.771	-0.4	0.9	18.6	76.1	12.5	40.2	32.2	100
TAG 56:9	+	918.7559	NA	1.5	NA	0	6.8	NA	6.5	NA	5
TAG 57:1	+	NA	948.8958	NA	0.5	2.3	19.3	2.4	5.5	0	3.6
TAG 57:2	+	NA	946.8817	NA	2.1	3.5	14.8	3	13.9	0.1	10.7
TAG 57:3	+	NA	944.8657	NA	1.7	10.5	15.9	6.6	22.5	4.4	33.8
TAG 57:4	+	942.8465	942.8483	-2	-0.1	9.3	27.3	4.6	11.6	3.3	18.7
TAG 57:5	+	NA	940.8322	NA	-0.6	2.3	26.1	2.7	13.3	0.7	15.6
TAG 57:6	+	938.8162	938.8188	-1	1.8	1.2	30.7	0.8	9.8	NA	14.2
TAG 57:7	+	936.8024	NA	1	NA	0	8	NA	4	NA	4.2
TAG 58:0	+	964.9267	NA	0.1	NA	0	14.8	NA	3	NA	3.5
TAG 58:10	+	944.7693	944.7702	-1	0	3.5	44.3	4.9	16.8	4.7	24.6
TAG 58:2	+	NA	960.8971	NA	1.8	19.8	44.3	54.4	32.1	178.7	54.5
TAG 58:3	+	958.882	NA	2.4	NA	1.2	18.2	9.8	4.4	NA	2.7
TAG 58:4	+	956.8661	NA	2.1	NA	1.2	19.3	4.8	4.9	NA	4.2
TAG 58:5	+	954.8483	954.8503	-0.1	2	3.5	73.9	2	26.5	1.3	54.2
TAG 58:6	+	952.8315	952.8336	-1.4	0.9	20.9	83	9.4	45.2	19.1	103.9
TAG 58:7	+	950.8156	950.8178	-1.6	0.8	25.6	87.5	11.3	44.5	31	107.2

continued on next page . . .

. . . continued from previous page

Species	Ion mode	*m/z* in sub-cohort		Error [ppm] in sub-cohort		Occupation [%]		Mean [pmol/mg tissue]		SD [pmol/mg tissue]	
		1	2	1	2	alveolar	tumor	alveolar	tumor	alveolar	tumor
TAG 58:8	+	948.802	948.8038	0.6	2.5	17.4	87.5	14.5	54.3	36.6	123.1
TAG 59:2	+	NA	974.911	NA	0	4.7	17	23.3	9.2	40.9	8.9
TAG 60:1	+	990.9436	NA	1.3	NA	0	17	NA	6.8	NA	7.6
TAG 60:10	+	972.8001	972.8004	-1.4	-1.1	3.5	55.7	5.7	22.9	5.2	39.2
TAG 60:12	+	968.7723	NA	2.2	NA	0	6.8	NA	3.3	NA	1.3
TAG 60:5	+	982.8822	NA	2.6	NA	0	13.6	NA	4.8	NA	4.4
TAG 60:6	+	980.8649	NA	0.9	NA	0	13.6	NA	5.8	NA	5
TAG 60:7	+	978.847	978.8499	-1.5	1.5	1.2	51.1	5	13.6	NA	30.6
TAG 60:8	+	976.8307	976.8335	-2.1	0.8	1.2	59.1	2.3	24.7	NA	47.5
TAG 60:9	+	974.8163	974.8181	-0.9	1	1.2	56.8	6.1	26.1	NA	53.4
TAG 62:13	+	994.7852	NA	-0.7	NA	0	6.8	NA	2.1	NA	1

B Index

Acronyms

A	alveolar tissue
AcOCl	acetyl chloride
ADC	adenocarcinoma
AECI	alveolar epithelial cells type I
AECII	alveolar epithelial cells type II
AM	alveolar macrophages
amu	atomic mass units
au	arbitrary units
AUC	area under the curve
BALF	brochio-alveolar lavage fluid
BHT	butylhydroxy toluene
BMI	body mass index
CCS	collisional cross-sections
CDP	cytidine diphosphate
CE	cholesteryl ester
Cer	ceramide
CHO_2^-	formate
CID	collision-induced dissociation
CL	cardiolipin
CoA	coenzyme A
COPD	chronic obstructive pulmonary disease
***.csv**	comma-separated-values

cv	cross-validation
D	deuterium, ^{2}H
DA	discriminant analysis
DAG	diacylglycerol
db	carbon-carbon double bond
DNA	deoxyribonucleic acid
dt	detection threshold
ESI	electrospray ionization
even numbered FA	fatty acid with even numbers of carbon atoms
FA	fatty acid
FC	free cholesterol
FDR	false discovery rate
FT-ICR	Fourier-transform ion cyclotron resonance
FWHM	full-width-at-half-maximum
$\times$**g**	times the g-force
GC	gas chromatography
GL	glycerolipid
GOLD	global initiative for chronic obstructive lung disease
GPL	glycerophospholipid
GSL	glyco-sphingolipid
Hb	hemoglobin
HE	hematoxylin and eosin
HexCer	hexosylceramide
HG	head group
HOPE	4-(5-hydroxylethyl)-1-piperazineethanesulfonic acid (HEPES)-glutamic acid buffer mediated organic solvent protection effect
HPLC	high pressure liquid chromatography
ID	identifier

IM	ion mobility
ISD	internal standard
LC	liquid chromatography
LCC	large-cell carcinoma
LCL	*lyso*-cardiolipin
LIFS	lipid*omics* informatics for life science
LPA	*lyso*-phosphatidic acid
LPC	*lyso*-phosphatidylcholine
LPE	*lyso*-phosphatidylethanolamine
LPG	*lyso*-phosphatidylglycerol
LPI	*lyso*-phosphatidylinositol
LPS	*lyso*-phosphatidylserine
LUX	lipidome juxtaposition
$[M+CHO_2]^-$	formate adduct ions
$[M+Cl]^-$	chlorinated adduct ions
$[M+H]^+$	protonated molecular ions
$[M-H]^-$	deprotonated molecular ions
$[M+NH_4]^+$	ammonium cation adduct ions
$[M+OAc]^-$	acetate adduct ions
Δm	mass difference
MAG	monoacylglycerol
MALDI	matrix assisted laser desorption/ionization
MFQL	molecular fragmentation query language
MS	mass spectrometry
MS^2	tandem mass spectrometry
MSI	mass spectrometric imaging
MTBE	*tert*-butyl methyl ether
m/z	mass-to-charge ratio

NET	neuroendocrine tumor
NH_4^+	ammonium
NH_4Cl	ammonium chloride
NH_4OAc	ammonium acetate
NLS	neutral-loss scanning
NSCLC	non-small-cell lung cancer
OAc^-	actetate
odd numbered FA	fatty acid with odd numbers of carbon atoms
OH	hydroxylation
OzID	ozone-induced-dissociation
PA	phosphatidic acid
PB	Paterno-Büchi reaction
PC	phosphatidylcholine
pc	principal component
pca	principal component analysis
PC *O-*	phosphatidylcholine ether
PCSNR	percentage change of signal-to-noise ratios
PE	phosphatidylethanolamine
PEEK	poly-ether-ether-ketone
PEG	polyethyleneglycol
PE *O-*	phosphatidylethanolamine ether
PG	phosphatidylglycerol
PI	phosphatidylinositol
PIS	precursor ion scan
PLS	partial least-squares
ppm	parts-per-million
PS	phosphatidylserine
PUFA	poly-unsaturated fatty acid

PY	pack years
Q	quadrupole
Q^2	cross-validated regression coefficient
QqQ	triple quadruple mass spectrometer
R^2	linear regression coefficient
RNA	ribonucleic acid
ROC	receiver operating characteristics
RP	resolving power
RSD	relative standard deviation
S1P	sphingosine-1-phosphate
SCA	sarcomatoid tumor
SCC	squamous-cell carcinoma
SD	standard deviation
SL	sphingolipid
SM	sphingomyelin
sn	stereospecific numbering
SNR	signal-to-noise ratio
SP	surfactant protein
surfactant	*surface-active agent*
t	components calculated for predictor variables
T	tumor tissue
TAG	triacylglycerol
TLC	thin-layer chromatography
TOF	time of flight
***.txt**	tab-delimited text
U	Wilcoxon Mann-Whitney U

List of Figures

List of Tables

C Personal Information

Curriculum Vitae

Publications

Research Articles

- LF Eggers, J Müller, C Marella, V Scholz, H Watz, C Kugler, KF Rabe, T Goldmann and D Schwudke, *Lipidomes of lung cancer and tumour-free lung tissues reveal distinct molecular signatures for cancer differentiation, age, inflammation, and pulmonary emphysema*, Scientific Reports, 7: 11087, **2017**, doi:10.1038/s41598-017-11339-1.
- D Hildebrand, P Merkel, LF Eggers, H Schlüter, *Proteolytic Processing of Angiotensin-I in Human Blood Plasma*, PLoS One, 8(5), e64027, **2013**, doi:10.1371/journal.pone.0064027.

Book Chapters

- LF Eggers and Dominik Schwudke, *Shotgun Lipidomics Approach for Clinical Samples*, Methods in Molecular Biology, Clinical Metabolomics: Methods and Protocols, Springer, **2018**, doi: 10.1007/978-1-4939-7592-1.
- LF Eggers and D Schwudke, *Lipid Extraction: Basics of the Methyl-tert-Butyl Ether Extraction*, M. Wenk (ed.), Encyclopedia of Lipidomics, Springer Science+Business Media Dordrecht **2016**, doi: 10.1007/978-94-007-7864-1_96-1.
- LF Eggers and D Schwudke, *Liquid Extraction: Folch*, M. Wenk (ed.), Encyclopedia of Lipidomics, Springer Science+Business Media Dordrecht **2016**, doi: 10.1007/978-94-007-7864-1_89-1.
- A Sündermann, LF Eggers and D Schwudke, *Liquid Extraction: Bligh and Dyer*, M. Wenk (ed.), Encyclopedia of Lipidomics, Springer Science+Business Media Dordrecht **2016**, doi: 10.1007/978-94-007-7864-1_88-1.

Project Presentations

Talks

- 50th Annual Conference of the German Society for Mass Spectrometry (DGMS), March 7, **2017**, Kiel, Germany, *Lipidomes of human lung tissues show distinct lipid metabolic alterations for lung cancer and pulmonary emphysema*.
- 49th Annual Conference of the German Society for Mass Spectrometry (DGMS), March 1, **2016**, Hamburg, Germany, *Lipidome studies of human lung tissues show metabolic alterations in lung cancer and emphysema*.
- 48th Annual Conference of the German Society for Mass Spectrometry (DGMS), March 2, **2015**, Wuppertal, Germany, *Lipidomics screens of human lung tissues together with histological characterization allow studying lipid metabolic perturbation in cancer and COPD*.

Posters

- 65th Conference on Mass Spectrometry and Allied Topics, American Society for Mass Spectrometry (ASMS), June 4-8, **2017**, Indianapolis, USA, *Lipid metabolic alterations hallmark pathological processes in human lungs*.
- 1st Lipidomics Forum, November 15-17, **2015**, Borstel, Germany, *The human lung lipidome: A potential connection of histopathological phenotypes and development of cancer or emphysema*.
- Annual Meeting of the German Center for Lung Research (DZL), January 26-27, **2015**, Hamburg, Germany, *Shotgun lipidomics and histology of human lung tissues enable insights into the pathology of COPD and cancer*.

Acknowledgments

My special thank goes to my project supervisor Dr. Dominik Schwudke who taught me shotgun lipid*omics*, critically reviewed my work and had always an open door for questions and discussion. Furthermore, I thank my co-supervisor Prof. Dr. Torsten Goldmann for providing me interesting insights into histopathology and cell biology, advise and critical discussion of my work.

There were many people who actively supported this project. Special thanks goes to Dr. Julia Müller who conceived the histopathological scoring method and spent many evenings at the microscope determining scores for each tissue sample. Next, I would like to acknowledge Verena Scholz who taught me lipid extractions and extracted a large number of the tissues investigated in this dissertation. I would like to thank Dr. Chakravarthy Marella for analyzing my data with his LUX score methodology. Thank goes also to Jasmin Tiebach and Maria Lammers for organizing the tissue samples in the pathology department and to the colleagues from the LungenClinic Großhansdorf, to PD Dr. Henrik Watz from who provided clinical data from the archive, and to Dr. Christian Kugler and his team for providing lung tissue samples after surgery. Thanks to Michael Weinkauf who introduced me to the Q-TOF mass spectrometer and established with me the quantification method for free cholesterol.

My tank goes to my former colleagues Dr. Matthias Krajewski, and Dr. Nicole Zehethofer who welcomed me in the research group and always answered my questions and discussed my results. I would like to thank Dr. Adam Wutkowski, Dr. Fadi Al Machot and Dr. Alena Woeste for proof-reading and critical commenting my dissertation.

Three years doctoral studies at the Research Center Borstel were a long time. Therefore, I send my thank to all my colleagues and collaboration partners I met during this time for a nice working atmosphere, nice meetings and discussions at conferences.

And last but not least, my special thank goes to my family for motivating me to study chemistry starting from first day until now. My parents supported me all the time and I am very grateful for this. My wife Julia and my now two-years-old son Max were the best motivation and support a doctoral student could have during this time, and sorry Max for, missing too many good night stories in the last months. Thank You!