

# **Large-Scale Clinical Platelet Lipidomics in Coronary Artery Disease Patients by Untargeted Data Independent Liquid Chromatography Hyphenated with High-Resolution Mass Spectrometry**

## **Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät  
der Eberhard Karls Universität Tübingen  
zur Erlangung des Grades eines  
Doktors der Naturwissenschaften  
(Dr. rer. nat.)

vorgelegt von  
Kristina Dittrich  
aus Kiel

Tübingen  
2023



Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der  
Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:

18.07.2023

Dekan:

Prof. Dr. Thilo Stehle

1. Berichterstatter:

Prof. Dr. Michael Lämmerhofer

2. Berichterstatter:

Prof. Dr. Stefan Laufer





# Contents

|                                                                     |            |
|---------------------------------------------------------------------|------------|
| <b>Summary</b>                                                      | <b>VI</b>  |
| <b>Zusammenfassung</b>                                              | <b>IX</b>  |
| <b>List of publications</b>                                         | <b>XII</b> |
| <b>Poster presentations</b>                                         | <b>XIV</b> |
| <b>Oral presentations</b>                                           | <b>XV</b>  |
| <b>Abbreviations</b>                                                | <b>XVI</b> |
| <b>1 Introduction</b>                                               | <b>1</b>   |
| 1.1 Platelets and their role in cardiovascular disease . . . . .    | 1          |
| 1.1.1 Structure and function of platelets . . . . .                 | 1          |
| 1.1.2 Haemostasis . . . . .                                         | 3          |
| 1.1.3 Cardiovascular disease . . . . .                              | 7          |
| 1.2 Lipids . . . . .                                                | 9          |
| 1.2.1 Lipid classification . . . . .                                | 9          |
| 1.2.2 Biology of lipids . . . . .                                   | 12         |
| 1.3 Lipidomics - Techniques and analytical strategies . . . . .     | 14         |
| 1.3.1 Sample preparation . . . . .                                  | 14         |
| 1.3.2 Liquid chromatography . . . . .                               | 15         |
| 1.3.2.1 Lipid species separation . . . . .                          | 15         |
| 1.3.2.2 Lipid class separation . . . . .                            | 16         |
| 1.3.3 Mass spectrometry . . . . .                                   | 17         |
| 1.3.3.1 Ionization . . . . .                                        | 17         |
| 1.3.3.2 Mass analyzer . . . . .                                     | 19         |
| 1.3.3.3 Detectors . . . . .                                         | 22         |
| 1.3.4 Data acquisition . . . . .                                    | 23         |
| 1.3.4.1 Targeted approach . . . . .                                 | 24         |
| 1.3.4.2 Untargeted approach with data dependent acquisition . . . . | 25         |
| 1.3.4.3 Untargeted approach with data independent acquisition . . . | 26         |
| 1.4 Large-scale lipidomics . . . . .                                | 27         |
| 1.4.1 Study and experimental design . . . . .                       | 28         |
| 1.4.2 Quality control . . . . .                                     | 30         |
| 1.4.3 Data processing strategy . . . . .                            | 31         |
| <b>2 Objective</b>                                                  | <b>37</b>  |

|          |                                                                               |           |
|----------|-------------------------------------------------------------------------------|-----------|
| <b>3</b> | <b>Material and methods</b>                                                   | <b>39</b> |
| 3.1      | Materials . . . . .                                                           | 39        |
| 3.2      | Methods . . . . .                                                             | 40        |
| 3.2.1    | Sample preparation and data acquisition . . . . .                             | 40        |
| 3.2.1.1  | Blood sampling . . . . .                                                      | 40        |
| 3.2.1.2  | Platelet isolation . . . . .                                                  | 41        |
| 3.2.1.3  | Lipid extraction . . . . .                                                    | 41        |
| 3.2.1.4  | Preparation of quality control samples . . . . .                              | 42        |
| 3.2.1.5  | Preparation of calibration samples . . . . .                                  | 42        |
| 3.2.1.6  | RP-UHPLC-ESI-MS/MS and sequence design . . . . .                              | 43        |
| 3.2.1.7  | Method validation . . . . .                                                   | 45        |
| 3.2.1.8  | Standard operating procedures (SOPs) . . . . .                                | 45        |
| 3.2.2    | Data processing and statistical analysis . . . . .                            | 46        |
| 3.2.2.1  | Quality control charts . . . . .                                              | 46        |
| 3.2.2.2  | Batch-wise untargeted data extraction . . . . .                               | 47        |
| 3.2.2.3  | Inter-batch feature alignment . . . . .                                       | 47        |
| 3.2.2.4  | Data curation of reference feature list . . . . .                             | 48        |
| 3.2.2.5  | Targeted data extraction and datamatrix preparation . . . . .                 | 49        |
| 3.2.2.6  | Optimization and application of data normalization . . . . .                  | 50        |
| 3.2.2.7  | Final datamatrix preparation . . . . .                                        | 51        |
| 3.2.2.8  | Statistical analysis . . . . .                                                | 52        |
| <b>4</b> | <b>Results and discussion</b>                                                 | <b>53</b> |
| 4.1      | Method validation . . . . .                                                   | 53        |
| 4.2      | Quality control . . . . .                                                     | 57        |
| 4.3      | Overview of the data processing workflow . . . . .                            | 62        |
| 4.4      | Inter-batch feature alignment . . . . .                                       | 64        |
| 4.4.1    | Tool design and application . . . . .                                         | 64        |
| 4.4.2    | Performance evaluation . . . . .                                              | 68        |
| 4.4.3    | Effect of alignment parameters on performance . . . . .                       | 71        |
| 4.4.4    | Improvement in lipid identification . . . . .                                 | 73        |
| 4.5      | Reference feature list preparation . . . . .                                  | 77        |
| 4.5.1    | Data curation of identified lipids . . . . .                                  | 78        |
| 4.5.2    | Data curation of unknown lipids . . . . .                                     | 86        |
| 4.5.3    | Final reference target list . . . . .                                         | 90        |
| 4.6      | Preparation of the large-scale lipidomics datamatrix . . . . .                | 95        |
| 4.6.1    | Targeted data extraction . . . . .                                            | 95        |
| 4.6.2    | Data normalization . . . . .                                                  | 100       |
| 4.6.3    | Final lipidomics datamatrix . . . . .                                         | 107       |
| 4.7      | Considerations for further methodological and workflow optimization . . . . . | 108       |

|                                                                                                                        |            |
|------------------------------------------------------------------------------------------------------------------------|------------|
| 4.8 Global platelet lipid profiling - Insights into cardiovascular disease, its therapy, and risk assessment . . . . . | 109        |
| 4.8.1 Significant lipidome alterations in dependence on disease severity . . .                                         | 110        |
| 4.8.2 Effects of statin treatment on the platelet lipidome . . . . .                                                   | 117        |
| 4.8.3 Cardiovascular risk assessment in CVD patients . . . . .                                                         | 121        |
| 4.8.4 Platelet-derived therapeutic targets - ACKR3/CXCR7 . . . . .                                                     | 126        |
| <b>5 Conclusion</b>                                                                                                    | <b>130</b> |
| <b>References</b>                                                                                                      | <b>133</b> |
| <b>A Appendix</b>                                                                                                      | <b>146</b> |
| A.1 Data processing parameters in MS-DIAL . . . . .                                                                    | 146        |
| A.2 Inter-batch feature alignment . . . . .                                                                            | 149        |
| A.3 Data curation of identified features . . . . .                                                                     | 149        |
| A.4 Targeted data extraction in MultiQuant . . . . .                                                                   | 152        |
| A.5 Lipidome alterations in dependence on disease severity . . . . .                                                   | 153        |
| A.6 Lipidome alterations in dependence on statin treatment . . . . .                                                   | 154        |
| A.7 Cardiovascular risk assessment in CAD . . . . .                                                                    | 155        |
| <b>List of Figures</b>                                                                                                 | <b>157</b> |
| <b>List of Tables</b>                                                                                                  | <b>160</b> |



## Summary

Technological advances and innovations in chromatography and mass spectrometry over the last decades promoted the comprehensive exploration of various molecules. The most prominent disciplines among the so-called "omics" approaches are proteomics and metabolomics. Among the latter, lipidomics is dedicated to the investigation of lipids, experiencing rising attention in biochemical and clinical research. Lipids are essential metabolites in constructing cellular structure, are key players in energy metabolism, and serve as intra- and intercellular signaling mediators. Further, lipids are crucially involved in the development, progression, and prognosis in many diseases, especially those of central health concern like cardiovascular diseases or cancer. Cardiovascular diseases are highly prevalent in modern societies with significant contributions on morbidity and mortality. Therefore, research in this field is expedited and here, lipidomics emerged to a valuable tool in assessing the physiological and pathophysiological lipid profile in human blood. The majority of studies are conducted on plasma or serum samples, however, the lipidome of platelets is of increasing interest. Platelets build the cellular compartment in haemostasis, but are also involved in development and progression of thrombosis, thrombo-inflammation, and immune response, all mechanisms regulated under lipid contributions. In this thesis, large-scale clinical lipidomics conducted on platelets from coronary artery disease (CAD) patients was performed to explore the lipid profile in dependence on various clinical conditions.

The lipidome analysis was performed in two CAD patient cohorts, one consisted of 120 patients assigned as prestudy cohort and the main study cohort included 1011 samples. The cohorts differed in the time of sampling (peri- vs. postinterventional) and sampling method (arterial vs. venous blood). Lipidomics analysis of both sample sets was based on reversed phase ultra-high-performance liquid chromatography (RP UHPLC) hyphenated with high resolution mass spectrometry on a hybrid quadrupole-time-of-flight (QTOF) instrument using untargeted data independent acquisition with sequential window acquisition of all theoretical fragment ion mass spectra (SWATH) for comprehensive MS and MS/MS data collection. Data were acquired in several batches, i.e. three and 22 in pre- and main study, respectively. The challenge of comprehensive lipid profiling data from large-scale ( $n > 100$ ) sample sets is the data processing, due to the data complexity having MS and MS/MS data from thousands of metabolites analyzed in several hundreds of samples in multiple-batch acquisitions. Thus, within the prestudy, a new data processing workflow was established, which was further refined during main study processing.

For the processing of untargeted lipidomics data, a combined untargeted-targeted data extraction strategy was proposed. Based on batch-wise untargeted data extraction, applied on a small number of batches, reference feature lists (feature = pair  $m/z$  und retention time) were generated for both cohorts serving as target list for further targeted data extraction. An inter-batch feature alignment tool to harmonize the feature lists from batch-wise untargeted data processing into a reference feature list was developed. The assessment of the tool performance

revealed good properties in the alignment of identical features in different batches at low levels of misaligned non-identical features (precision: 98.5%, recall: 99.8%). Further, a significant improvement in lipid identification was achieved when considering multiple batches for the reference target list generation.

To obtain high quality lipidomics datamatrices, data curation of the lipid annotations but also of detected features with unknown identity was mandatory to identify and remove the feature targets being misannotations or artefact peaks. Through manual and software-assisted curation of identified and, in the main study, unknown features, the target lists were significantly reduced to valid features. Further, in the main study, a harmonization of lipids detected in both acquired polarity modes was performed to remove redundant lipid entries in the target list. The reference target lists of both studies were subjected to targeted data extraction. To account for unwanted variations in LC-MS-MS/MS signal responses, introduced by instrumental and inter-day fluctuations arising in multi-batch large-scale studies, data normalization was applied after method evaluation and selection of the optimum normalization method. After removal of features exhibiting coefficient of variation larger 30% in normalized responses of regularly implemented quality control measurements, the prestudy lipidomics datamatrix included 3915 features, among which 939 (24%) were annotated lipids. In the main study, the datamatrix contained 4013 features with 767 identified lipids (19%).

The platelet lipid profiles of the prestudy cohort were evaluated for lipidome alterations in dependence on disease severity and statin treatment. When comparing the platelet lipidome of patients with chronic coronary syndrome (CCS) against the more severe state of acute coronary syndrome (ACS), significant alterations were observed involving the upregulation of medium chain fatty acyl phosphatidylcholines (MCFA-PCs) and the downregulation of polyunsaturated (PUFA) phospholipids in ACS patients. The increase of MCFA-PCs levels seems to lead to platelet hyperreactivity and was assumed to originate from a fatty acid oxidation disorder in mitochondria and/or peroxisomes. The assessment of the statin effect on the platelet lipidome indicated a softening modulation of the lipidome as a pleiotropic effect of lipid-lowering therapy, associated with an increase of energy storage related lipids triacylglycerides (TGs) and PUFA cholesteryl esters (PUFA-CEs). Accompanied with an observed decrease of disease-related MCFA-PCs upon statin treatment, statins might beneficially modulate lipid metabolism, i.e.  $\beta$ -oxidation and synthesis. Furthermore, lipidomics data from the prestudy supported experimental studies highlighting the strength of ACKR3/CXCR7 agonist as a new therapeutic strategy in CAD treatment in contrast to conventional antiplatelet therapy by acetylsalicylic acid, as administered to a subset of patients from the clinical CAD cohort.

Based on the main study cohort of more than 1000 CAD patients, lipid profile associations with a risk of adverse cardiovascular (CV) events within a three-year follow-up were explored. Differentiating between ischemic and bleeding events, lipidome alterations were observed involving predominantly upregulation of lysophosphatidylethanolamines (LPEs) in ischemic events and increased levels of acylcarnitine (ACar) and fatty acid (FA) species in

bleeding events, respectively, compared to eventless patients. Further, biomarker evaluation for risk stratification was performed, identifying a small lipid panel which was able to improve the CV risk prediction in combination with conventional risk factors.

## Zusammenfassung

Technologische Fortschritte und Innovationen der letzten Jahrzehnte in den Bereichen Chromatographie und Massenspektrometrie haben die umfassende Untersuchung verschiedenster Moleküle vorangetrieben. Die bekanntesten Disziplinen unter den als "Omik" bezeichneten Methoden sind Proteomik und Metabolomik. Zu letzterer gehörend ist die Lipidomik der Untersuchung von Lipiden gewidmet und erfährt wachsende Bedeutung in der biochemischen und klinischen Forschung. Lipide sind essentielle Metabolite, die an der Ausbildung zellulärer Strukturen beteiligt sind, eine Schlüsselrolle im Energiestoffwechsel spielen und als intra- und interzelluläre Signalmoleküle dienen. Weiterhin sind Lipide entscheidend beteiligt an der Ausbildung, dem Fortschreiten und der Prognose vieler Erkrankungen, insbesondere jenen mit zentraler gesundheitlicher Bedeutung wie kardiovaskuläre Erkrankungen oder Krebs. Kardiovaskuläre Erkrankungen sind sehr prevalent in modernen Gesellschaften und haben einen signifikanten Anteil an Morbidität und Mortalität. Daher wird die Forschung in diesem Bereich vorangetrieben und die Lipidomik hat sich hierbei zu einer wertvollen Methode zur Untersuchung des physiologischen und pathophysiologischen Lipidprofils in humanem Blut entwickelt. Der Großteil der Studien basiert auf der Untersuchung von Plasma oder Serum, allerdings steigt das Interesse am Lipidom der Blutplättchen. Blutplättchen bilden den zellulären Bestandteil der Haemostase, sind aber auch an der Entwicklung und dem Voranschreiten von Thrombose, Thrombo-Inflammation und Immunantwort beteiligt. Lipide sind in die Regulation all dieser Prozesse involviert. Im Rahmen dieser Arbeit wurde eine umfangreiche klinische Lipidomanalyse mit Blutplättchen von Patienten mit koronarer Herzkrankheit (KHK) durchgeführt, um das Lipidprofil in Abhängigkeit verschiedener klinischer Parameter zu untersuchen.

Die Lipidomanalyse wurde mit zwei Patientenkohorten mit KHK durchgeführt, eine, als Vorstudie bezeichnet, bestand aus 120 Patienten und die Hauptstudie umfasste 1011 Proben. Die Kohorten unterschieden sich im Zeitpunkt der Probennahme (peri- vs. postinterventionell) und in der Entnahmemethodik (arterielles vs. venöses Blut). Die Lipidomik-Messungen beider Probensets erfolgte mittels Umkehrphasen-Hochleistungsflüssigchromatographie (RP UHPLC) verbunden mit hochauflösender Massenspektrometrie mit einem Hybrid Quadrupol-Flugzeit-Massenspektrometer (QTOF) in ungezielter (engl. *untargeted*) daten-unabhängiger Akquisition mittels SWATH-Technik (engl. *sequential window acquisition of all theoretical fragment ion mass spectra*) für eine globale Erfassung von MS und MS/MS Daten. Die Messungen erfolgten in mehreren Durchläufen, d.h. Batches, drei in der Vor- und 22 in der Hauptstudie. Die Herausforderung globaler Lipidprofil-Daten einer umfangreichen Kohorte (>100 Proben) liegt in der Datenprozessierung, aufgrund der Komplexität der Daten mit MS und MS/MS Spektren von tausenden Metaboliten aus mehreren hundert Proben erhoben in mehreren Messdurchläufen. Daher wurde im Rahmen der Vorstudie ein neuer Arbeitsablauf zur Datenprozessierung etabliert, der im Verlauf der Hauptstudie weiterentwickelt wurde.

Zur Prozessierung von *untargeted* erhobenen Lipidomik-Daten wurde ein kombinierter



untargeted-targeted Ansatz eingeführt. Basierend auf einer ungerichteten Datenextraktion einzelner Batches wurden Referenzlisten an Features (Paar aus  $m/z$  und Retentionszeit) für jede Kohorte erstellt, die im weiteren Verlauf zur gezielten Datenextraktion verwendet wurde. Ein Programm zur Angleichung der Features zwischen den Batches (engl. inter-batch feature alignment tool) wurde entwickelt, um aus Batch-spezifischen Feature Listen eine Referenzliste zu erzeugen. Die Leistungsüberprüfung des Programms zeigte eine gute Fähigkeit zur Angleichung identischer Features aus unterschiedlichen Batches mit einem geringen Anteil an fälschlich zusammengeführten, nicht identischen Features (Präzision: 98.5%, Trefferquote: 99.8%). Darüber hinaus konnte eine signifikante Verbesserung der Lipididentifizierung erzielt werden, indem mehrere Batches zur Erzeugung der Referenzliste herangezogen wurden.

Um qualitativ hochwertige Lipidom-Datenmatrizes zu erhalten, ist das Kuratieren der Lipididentifizierungen aber auch von Features unbekannter Identität unerlässlich, um Features mit falscher Lipidnotation oder Peakartefakte zu identifizieren und zu entfernen. Durch manuelle und Software-unterstützte Kuratierung identifizierter und, in Bezug auf die Hauptstudie, unbekannter Features wurde die Target-Liste signifikant auf valide Features reduziert. Zusätzlich wurden in der Hauptstudie Features harmonisiert, die in beiden gemessenen Polaritäten detektiert wurden, und nur in der Target-Liste des bevorzugten Polaritätsmodus beibehalten. Die Target-Referenzlisten beider Studien wurden anschließend zur gezielten (targeted) Datenextrahierung eingesetzt. Zur Berücksichtigung von ungewollt auftretender Variation in den LC-MS-MS/MS Signalen, die durch instrumentelle und tagesabhängige Fluktuationen zustande kommen, insbesondere in umfangreichen Studien mit vielen Batches, wurde eine Normalisierung der Daten durchgeführt, nachdem verschiedene Methoden evaluiert und die optimale Methode ausgewählt wurde. Nachdem Features mit einem Variationskoeffizienten größer 30% in normalisierten Qualitätskontrollproben, die regelmäßig verteilt über beide Studien gemessen wurden, entfernt wurden, enthielt die Lipidomik-Datenmatrix der Vorstudie 3915 Features mit 939 identifizierten Lipiden (24%) und die Hauptstudie 4013 Features mit 767 identifizierten Lipiden (19%).

Die Lipidprofile der Blutplättchen aus der Vorstudien-Kohorte wurden auf Lipidomveränderungen in Abhängigkeit von der Schwere der KHK-Erkrankung und einer Statinbehandlung untersucht. Beim Vergleich der Plättchenlipidome von Patienten mit chronischem Koronarsyndrom (CCS) mit dem schwereren Zustand des akuten Koronarsyndroms (ACS) haben sich signifikante Unterschiede herauskristallisiert, wobei Phosphatidylcholine mit mittelkettigen Fettsäureseitenketten (MCFA-PCs) herauf- und mehrfach ungesättigte (PUFA) Phospholipide herunterreguliert waren. Die Zunahme der MCFA-PC Level scheint zur Hyperreaktivität der Blutplättchen beizutragen und es wurde vermutet, dass dieser Anstieg durch eine Störung des mitochondrialen und/oder peroxisomalen Fettsäurestoffwechsels hervorgerufen wird. Die Untersuchung des Einflusses einer Statinbehandlung auf das Plättchenlipidom wies auf eine mildernde Modulation des Lipidoms im Rahmen eines pleiotropen Effekts der Lipid-senkenden Therapie hin, die mit dem Anstieg von Triacylglyceriden (TGs) und PUFA Cholesterylestern (PUFA-CEs), beides Lipidklassen zur intrazellulären Energiespeicherung,

einherging. Zusammen mit einem beobachteten Rückgang der krankheits-assoziierten MCFA-PCs unter Statinbehandlung scheinen Statine den Lipidmetabolismus in Bezug auf  $\beta$ -Oxidation und Lipidsynthese positiv zu beeinflussen. Darüber hinaus wurden die Lipidomdaten der Vorstudie zur Unterstützung experimenteller Studien eingesetzt, die den Vorteil von ACKR3/CXCR7 Agonist als neuen Therapieansatz bei KHK zeigen, auch im Vergleich zur konventionellen Therapie mittels Acetylsalicylsäure, wie sie bei einem Teil der Kohortenpatienten angewendet wurde.

Basierend auf der Hauptstudienkohorte mit mehr als 1000 Patienten wurden die Lipidprofile im Hinblick auf Assoziationen mit dem Risiko für kardiovaskuläre (CV) Ereignisse im Zeitraum von drei Jahren nach Studienbeginn untersucht. Eine Unterscheidung zwischen ischämischen und Blutungsereignissen zeigte eventabhängige Lipidomveränderungen mit Heraufregulation von Lysophosphatidylethanolaminen (LPEs) in ischämischen Ereignissen und einer Zunahme der Level von Acylcarnitin-(ACar) und Fettsäurespezies (FA) bei Blutungsereignissen im Vergleich zu Patienten ohne kardiovaskuläres Ereignis. Weiterhin wurden Biomarker zur Risikostratifikation evaluiert, wobei ein kleines Panel an Lipiden identifiziert wurde, das in Kombination mit konventionellen Risikofaktoren zu einer Verbesserung der Risikovorhersage für kardiovaskuläre Ereignisse beitragen konnte.

## List of publications

### Clinical lipidomics related publications (thesis-related)

#### **Publication I**

Harm, T.; Dittrich, K.; Brun, A.; Fu, X.; Frey, M.; Uribe, A.P.; Schwarz, F.; Rohlfing, A.; Castor, T.; Geisler, T.; Rath, D.; Lämmerhofer, M.; Gawaz, M. Large-scale lipidomics profiling reveals characteristic lipid signatures associated with an increased cardiovascular risk. *Clin Res Cardiol* **2023**, DOI: 10.1007/s00392-023-02260-x.

#### **Publication II**

Harm, T.; Frey, M.; Dittrich, K.; Goldschmied, A.; Rohlfing, A.; Fu, X.; Brun, A.; Castor, T.; Rath, D.; Müller, K.; Lämmerhofer, M.; Gawaz, M. Statin treatment is associated with alterations in the platelet lipidome. *Thromb Haemost* **2023**, DOI: 10.1055/s-0043-1764353.

#### **Publication III**

Harm, T.; Bild, A.; Dittrich, K.; Goldschmied, A.; Nestele, J.; Chatterjee, M.; Fu, X.; Kolb, K.; Castor, T.; Borst, O.; Geisler, T.; Rath, D.; Lämmerhofer, M.; Gawaz, M. Acute coronary syndrome is associated with a substantial change in the platelet lipidome. *Cardiovasc Res* **2021**, 118(8):1904-1916, DOI: 10.1093/cvr/cvab238.

#### **Publication IV**

Cebo, M.; Dittrich, K.; Fu, X.; Manke, M.C.; Emschermann, F.; Rheinlaender, J.; von Eysmondt, H.; Ferreirós, N.; Sudman, J.; Witte, A.; Pelzl, L.; Borst, O.; Geisler, T.; Rath, D.; Bakchoul, T.; Gawaz, M.; Schäffer, T.E.; Lämmerhofer, M.; Chatterjee, M. Platelet ACKR3/CXCR7 favors antiplatelet lipids over an atherothrombotic lipidome and regulates thromboinflammation. *Blood* **2022**, 139(11):1722-1742, DOI: 10.1182/blood.2021013097.

#### **Publication V**

Dittrich, K.; Cebo, M.; Drotleff, B.; Calderón Castro, C.; Fu, X.; Lämmerhofer, M. LC-MS Lipidomik unter Verwendung von SWATH: Globale daten-unabhängige Erzeugung von Massenspektren. *Wiley Analytical Science* **2021**, 65(9):22-24, [analyticalscience.wiley.com/do/10.1002/was.00170254](https://doi.org/10.1002/was.00170254)

Publications regarding chiral stationary phases for enantioselective metabolomics and lipidomics (not associated with the thesis project but published during research period)

#### **Publication VI**

Schmitt, K.; Woiwode, U.; Kohout, M.; Zhang, T.; Lindner, W.; Lämmerhofer, M. Comparison of small size fully porous particles and superficially porous particles of chiral anion-exchange type stationary phases in ultra-high performance liquid chromatography: Effect of particle and pore size on chromatographic efficiency and kinetic performance. *J Chromatogr A* **2018**, 1569:149-159, DOI: 10.1016/j.chroma.2018.07.056

#### **Publication VII**

Geibel, C.; Dittrich, K.; Woiwode, U.; Kohout, M.; Zhang, T.; Lindner, W.; Lämmerhofer, M. Evaluation of superficially porous particle based zwitterionic chiral ion exchangers against fully porous particle benchmarks for enantioselective ultra-high performance liquid chromatography. *J Chromatogr A* **2019**, 1603:130-140, DOI: 10.1016/j.chroma.2019.06.026

#### **Publication VIII**

Schmitt, K.; Lämmerhofer, M. Optimization of surface modification process of crosslinked polythiol-coated chiral stationary phases synthesized by two-step thiol-ene click reaction. *J Sep Sci* **2018**, 41:1338-1345, DOI: 10.1002/jssc.201701425

#### **Publication IX**

Geibel, C.; Dittrich, K.; Wolter, M.; Lämmerhofer, M. Thiol-ene photo-click immobilization of a chiral chromatographic ligand on silica particles. *J Chromatogr A* **2020**, 1622:461133, DOI: 10.1016/j.chroma.2020.461133

## Poster presentations

**TR240 Retreat 2019**, Neckarsulm, Germany, May 5<sup>th</sup> - 7<sup>th</sup>.

Perspective of the platelet lipidome for diagnostic and therapeutic strategies.

Kristina Dittrich, Michael Lämmerhofer

**Annual DPhG-Meeting 2021**, Virtual Meeting, September 29<sup>th</sup> - October 1<sup>st</sup>.

Untargeted data-independent lipidomics of platelets in cardiovascular disease.

Kristina Dittrich, Tobias Harm, Andreas Goldschmied, Madhumita Chatterjee, Meinrad Gawaz, Michael Lämmerhofer

**4<sup>th</sup> International Symposium "Platelets" 2022**, Würzburg, Germany, June 23<sup>rd</sup> - 25<sup>th</sup>.

Platelet lipidomics profiling in a large-scale clinical cohort of CAD patients.

Kristina Dittrich, Xiaoqing Fu, Adrian Brun, Meinrad Gawaz, Tobias Geisler, Madhumita Chatterjee, Michael Lämmerhofer

## Oral presentations

**TR240 Retreat 2019**, Neckarsulm, Germany, May 5<sup>th</sup> - 7<sup>th</sup>.

Regulation of the platelet lipidome and its implications in thrombo-inflammation.

Kristina Dittrich, Hendrik von Eysmondt, Madhumita Chatterjee, Michael Lämmerhofer, Tilmann E. Schäffer

**TR240 Retreat 2020**, Virtual Meeting, September 29<sup>th</sup> - 30<sup>th</sup>.

Targeting chemokine receptor CXCR7 favours anti-platelet lipids over a pro-thrombotic lipidome and regulates thrombo-inflammation: potential implications for anti-thrombotic strategy in CAD.

Madhumita Chatterjee, Michael Lämmerhofer, Tilmann E. Schäffer, Kristina Dittrich, Hendrik von Eysmondt

**TR240 Seminar Series**, Virtual Meeting, Dezember 1<sup>st</sup>.

Platelet lipidomics.

Kristina Dittrich, Małgorzata Cebo, Xiaoqing Fu, Peng Li, Madhumita Chatterjee, Michael Lämmerhofer, Tilmann E. Schäffer

**Annual DPhG-Meeting 2021**, Virtual Meeting, September 29<sup>th</sup> - October 1<sup>st</sup>.

Untargeted data-independent lipidomics of platelets in cardiovascular disease.

Kristina Dittrich, Tobias Harm, Andreas Goldschmied, Madhumita Chatterjee, Meinrad Gawaz, Michael Lämmerhofer

## Abbreviations

|        |                                                                 |
|--------|-----------------------------------------------------------------|
| AA     | arachidonic acid                                                |
| AC     | adenylate cyclase                                               |
| ACN    | acetonitrile                                                    |
| ADP    | adenosine diphosphate                                           |
| APCI   | atmospheric pressure chemical ionisation                        |
| APT    | antiplatelet therapy                                            |
| ATP    | adenosine triphosphate                                          |
| B-MIS  | best-matched internal standard normalization                    |
| CAD    | coronary artery disease                                         |
| cAMP   | cyclic adenosine monophosphate                                  |
| CAR    | acylcarnitine                                                   |
| CCL5   | chemokine ligand 5                                              |
| CCMN   | cross-contribution compensating multiple standard normalization |
| CCS    | collision cross section                                         |
| CE     | cholesteryl ester                                               |
| CE     | collision energy                                                |
| Cer    | ceramide                                                        |
| CES    | collision energy spread                                         |
| cGMP   | cyclic guanosine monophosphate                                  |
| CID    | collision-induced dissociation                                  |
| CL     | cardiolipin                                                     |
| COX    | cyclooxygenase                                                  |
| CUR    | curtain gas                                                     |
| CV     | coefficient of variation                                        |
| CXCL   | chemokine                                                       |
| CXCR   | chemokine receptor                                              |
| CYP450 | cytochrome P450                                                 |
| DB     | double bond                                                     |

---

|                     |                                               |
|---------------------|-----------------------------------------------|
| DC                  | direct currecnt                               |
| DDA                 | data dependent acquisition                    |
| DG                  | diacylglycerol                                |
| DGLA                | dihomo- $\gamma$ -linolenic acid              |
| DIA                 | data independent acquisition                  |
| DP                  | declustering potential                        |
| EIC                 | extracted ion chromatogram                    |
| ESI                 | electrospray ionisation                       |
| FA                  | fatty acid                                    |
| FA                  | formic acid                                   |
| FAME                | fatty acid methyl ester                       |
| FAS                 | fatty acid synthase                           |
| FC                  | fold change                                   |
| FDA                 | U.S. Food and Drug Administration             |
| FDR                 | false discovery rate                          |
| FWHM                | full width at half maximum                    |
| GC                  | gas chromatography                            |
| GP                  | glycoprotein                                  |
| HAc                 | acetic acid                                   |
| Hex2Cer             | dihexosylceramide                             |
| HILIC               | hydrophilic interaction liquid chromatography |
| HR                  | hazard ratio                                  |
| HR                  | high resolution                               |
| ICAM-1              | intercellular adhesion molecule-1             |
| IDA                 | information dependent acquisition             |
| IPA                 | isopropanol                                   |
| IL-1 $\beta$        | interleukin 1 $\beta$                         |
| IP                  | prostacyclin receptor                         |
| IP <sub>3</sub> (R) | inositol-1,4,5-triphosphate (receptor)        |



|            |                                                  |
|------------|--------------------------------------------------|
| IS         | internal standard                                |
| ISVF       | ion spray voltage floating                       |
| ITC        | ion transmission control                         |
| LC         | liquid chromatography                            |
| LCFA       | long chain fatty acids                           |
| LIPID MAPS | Lipid Metabolites and Pathways Strategy          |
| LOD        | limit of detection                               |
| LOESS      | locally estimated scatterplot smoothing          |
| LOQ        | limit of quantification                          |
| LOWESS     | locally weighted estimated scatterplot smoothing |
| LOX        | lipoxygenase                                     |
| LPC        | lysophosphatidylcholine                          |
| LPC-O      | ether-linked lysophosphatidylcholine             |
| LPE        | lysophosphatidylethanolamine                     |
| LPE-O      | ether-linked lysophosphatidylethanolamine        |
| LPI        | lysophosphatidylinositol                         |
| LPS        | lysophosphatidylserine                           |
| LR         | low resolution                                   |
| LSM        | lysosphingomyelin                                |
| MAD        | median absolute deviation                        |
| MCFA(-PC)  | medium chain fatty acyl(-phosphatidylcholine)    |
| MCP        | microchannel plate                               |
| ME         | matrix effect                                    |
| MeOH       | methanol                                         |
| MG         | monoacylglycerol                                 |
| MRM        | multiple reaction monitorin                      |
| mRNA       | messengerRNA                                     |
| MS         | mass spectrometry                                |
| MS/MS      | tandem mass spectrometry                         |

|           |                                                                      |
|-----------|----------------------------------------------------------------------|
| mTIC      | total ion current with all annotated metabolites                     |
| MVDA      | multivariate data analysis                                           |
| $m/z$     | mass-to-charge                                                       |
| NAE       | N-acyl ethanolamine                                                  |
| NO        | nitric oxide                                                         |
| NOMIS     | normalization using optimal selection of multiple internal standards |
| NPLC      | normal phase liquid chromatography                                   |
| OCS       | open canalicular system                                              |
| OPLS(-DA) | orthogonal partial least squares (discriminant analysis)             |
| oxLDL     | oxidized low-density lipoprotein                                     |
| PAF       | platelet activating factor                                           |
| PAR       | protease-activated receptor                                          |
| PC        | phosphatidylcholine                                                  |
| PCA       | principal component analysis                                         |
| PCI       | percutaneous coronary intervention                                   |
| PCnorm    | normalization based on phosphatidylcholines                          |
| PC-O      | ether-linked phosphatidylcholine                                     |
| PE        | process efficiency                                                   |
| PE        | phosphatidylethanolamine                                             |
| PE-O      | ether-linked phosphatidylethanolamine (alkyl-linkage)                |
| PE-P      | ether-linked phosphatidylethanolamine (alkenyl-linkage)              |
| PF4       | platelet factor 4                                                    |
| PG        | phosphatidylglycerol                                                 |
| PG        | prostaglandin                                                        |
| PI        | phosphatidylinositol                                                 |
| PI3K      | phosphoinositide 3-kinases                                           |
| PKC       | protein kinase C                                                     |
| PL        | phospholipase                                                        |
| PLS(-DA)  | partial least squares (discriminant analysis)                        |

|           |                                              |
|-----------|----------------------------------------------|
| PPAR      | peroxisome proliferator-activated receptor   |
| ppm       | parts per million                            |
| PQN       | probabilistic quotient normalization         |
| PRM       | parallel reaction monitoring                 |
| PS        | phosphatidylserine                           |
| P-Sel     | P-selectin                                   |
| PSGL-1    | P-selectin glycoprotein ligand-1             |
| PUFA      | polyunsaturated fatty acid                   |
| QC        | quality control                              |
| QqQ       | triple quadrupole                            |
| QTOF      | quadrupole-time-of-flight                    |
| QTrap     | quadrupole linear ion trap                   |
| RDM       | research data management                     |
| RE        | recovery                                     |
| RF        | radio frequency                              |
| RF        | random forest                                |
| ROS       | reactive oxygen species                      |
| RPLC      | reversed phase liquid chromatography         |
| RT        | retention time                               |
| RUVrandom | remove unwanted variation random             |
| SCFA      | short chain fatty acids                      |
| SERRF     | systematic error removal using random forest |
| SFC       | supercritical fluid chromatography           |
| sGC       | soluble guanylyl cyclase                     |
| SHexCer   | sulfatide                                    |
| SIM       | selected ion monitoring                      |
| SM        | sphingomyelin                                |
| S/N       | signal-to-noise ratio                        |
| SRM       | selected reaction monitoring                 |

|       |                                                                            |
|-------|----------------------------------------------------------------------------|
| ST    | sterol lipid                                                               |
| STIM1 | stromal interaction molecule 1                                             |
| SVM   | support vector machines                                                    |
| SWATH | sequential window acquisition of all theoretical fragment ion mass spectra |
| TDC   | time-to-digital converter                                                  |
| TEM   | temperature                                                                |
| TG    | triacylglycerol                                                            |
| TIC   | total ion current                                                          |
| TP    | thromboxane receptor                                                       |
| TXA   | thromboxane                                                                |
| UHPLC | ultra-high-performance liquid chromatography                               |
| UVDA  | univariate data analysis                                                   |
| Var   | variance                                                                   |
| VBA   | Visual Basic for Applications                                              |
| VLCFA | very long chain fatty acids                                                |
| vWF   | von Willebrand factor                                                      |

# 1 Introduction

## 1.1 Platelets and their role in cardiovascular disease

Platelets are small, circulating blood cells being involved in manifold physiological and pathophysiological processes like haemostasis [1], thrombosis [2], and thrombo-inflammation [3, 4]. The following sections are dedicated to the platelet structure and their role in haemostasis and cardiovascular disease.

### 1.1.1 Structure and function of platelets

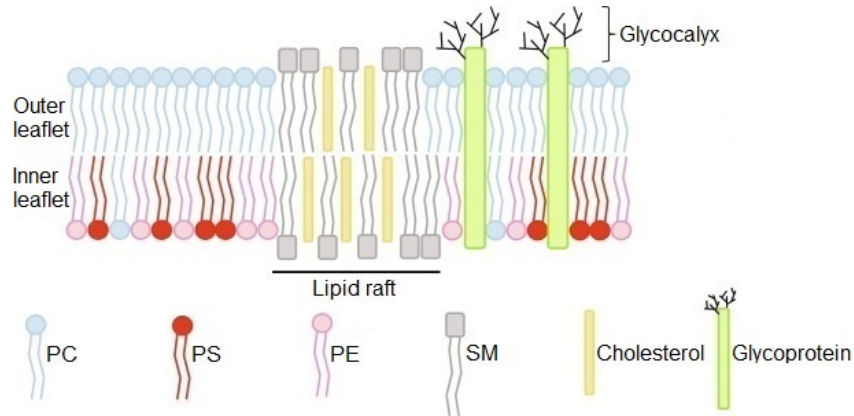
Platelets are small cells with a diameter of 2-4  $\mu\text{m}$  without a cell nucleus [1]. Due to their anucleated character, platelets are incapable of cell mitosis and limited in protein synthesis abilities [5, 6]. They have a short lifespan of 7-10 days [7].

Platelets are produced primarily in the bone marrow by thrombopoiesis. Haematopoietic stem cells differentiate to polyploid megakaryocytes, from which proplatelets are formed by cytoplasmatic protrusions. Platelets generation from megakaryocytes is under control of thrombopoietin [1, 8]. In human blood, a platelet count of 150 000 to 400 000 cells per microliter is maintained [7]. Several ways of platelet clearance are known, e.g. elimination of aged platelets in liver and spleen or consumption due to platelet function in haemostasis or thrombosis [1, 9].

Platelets have an asymmetric plasma membrane which is a lipid bilayer primarily composed of phospholipids phosphatidylcholine (PC), phosphatidylserine (PS), phosphatidylethanolamine (PE), the sphingolipid sphingomyelin (SM), and the sterol lipid cholesterol (Figure 1.1 A). While PCs are the major lipid class in the outer leaflet of the membrane, PS and PE are building the inner leaflet [5]. Lipid rafts are areas within the membrane being enriched with SM and cholesterol and are specialized signaling areas. Upon activation, the plasma membrane is remodeled and bioactive lipid species like fatty acids (FA), eicosanoids, prostaglandins, 1,2-diacylglycerol, and lysophospholipids are produced by enzymatic activities. Phospholipids are also the major building block of organelles membranes [10]. Neutral lipids, i.e. di- und triacylglycerols, as well as cholesteryl esters (CE) are stored intracellularly in lipid droplets [11].

Platelets are secretory cells, thus, besides mitochondria, platelets contain  $\alpha$ -granules, dense granules, lysosomes, and peroxisomes (Figure 1.1 B), from which bioactive substances are released stimulus-dependent [12, 13].

A



B

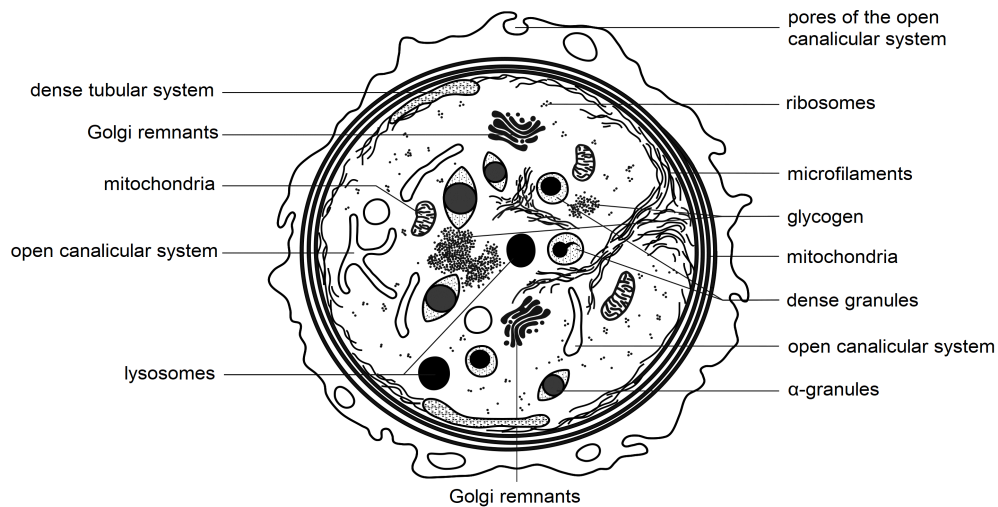


Figure 1.1: (A) The plasma membrane of platelets is an asymmetric lipid bilayer. Modified with permission from [5] (Copyright 2019, Elsevier). (B) Platelets cellular structure showing the cell organelles. Reprinted with permission from [14] (Copyright 2015, Neumüller, Ellinger and Wagner). Although not shown in this scheme, platelets also contain small peroxisomes [15].

Comprising 10% of the platelets cell volume, the  $\alpha$ -granules represent the major type of granules. In 50 to 80  $\alpha$ -granules per cell, membrane and adhesion proteins (e.g. fibrinogen, von Willebrand factor (vWF)), fibrinolytic proteins (e.g. plasminogen), coagulants, anticoagulants, chemokines (e.g. CXCL4, CXCL12), growth and angiogenic factors (e.g. EGF, PDGF), as well as immune mediators (e.g. IgG, complement precursors) are stored [12].

In contrast to  $\alpha$ -granules, dense granules are smaller and less abundant in platelets (3 to 8 per cell, 1% of the cell volume). Their cargos are crucial in haemostasis, since they contain adenosine diphosphate (ADP) and  $\text{Ca}^{2+}$  as auto- and paracrine actors in blood clotting. Moreover, dense granules carry other cations (e.g.  $\text{Mg}^{2+}$ ) and nucleotides (e.g. adenosine triphosphate (ATP), polyphosphates, and bioactive amines (e.g. serotonin and histamine) [12].

Besides their protein degradation function based on intraorganelle acidic conditions and hydrolytic enzymes, lysosomes also have secretory properties [16]. Lysosome contents are released stimulus-related during platelet activation, however, a greater stimulus is required compared to  $\alpha$ - and dense granules. It is assumed that lysosome released compounds are involved in thrombus formation regulation and extracellular matrix remodeling [16].

Finally, platelets have small peroxisomes, being responsible for fatty acid  $\beta$ -oxidation of very long chain fatty acids ( $> 22$  carbon atoms). They are oxidative organelles due to their catalase activity [15].

The open canalicular system (OCS) is an internal membrane network formed by channels of varying length and diameter providing connections to the outer plasma membrane. This channel complex is essential in the transport of metabolites from in- and outside of the platelet, e.g. releasing granules cargo upon activation. Furthermore, the open canalicular system is used as an intracellular reservoir of membrane which is required during shape remodeling under platelet activation [17].

In addition to the OCS, the dense tubular system (DTS) is a second intracellular but closed-channel system. In membrane-like tubules,  $\text{Ca}^{2+}$  is stored in high concentrations and released during platelet activation [14, 18].

Although lacking a nucleus, platelets are still able to perform protein synthesis [19]. Platelets contain the respective cellular equipment, i.e. endoplasmatic reticulum and ribosomes, as well as messengerRNA (mRNA), which originates from the megakaryocytes during thrombopoiesis [19–21]. Furthermore, microRNA is present in platelets enabling alterations in protein expression. RNA levels are decreasing with aging, thus RNA concentrations are higher in younger platelets which are called reticulated platelets [1]. Moreover, platelets contain nuclear receptors like peroxisome proliferator-activated receptors (PPAR), sex steroids and glucocorticoid receptors which contribute to platelet function modulation in a non-genomic manner [6].

### 1.1.2 Haemostasis

The primary physiological role of platelets is the clot formation to stop bleeding caused by vascular injury. Due to this, platelets contribute to a complex process called haemostasis. During haemostasis, the cell morphology is changed from a discoid to a spread cell shape (Figure 1.2).

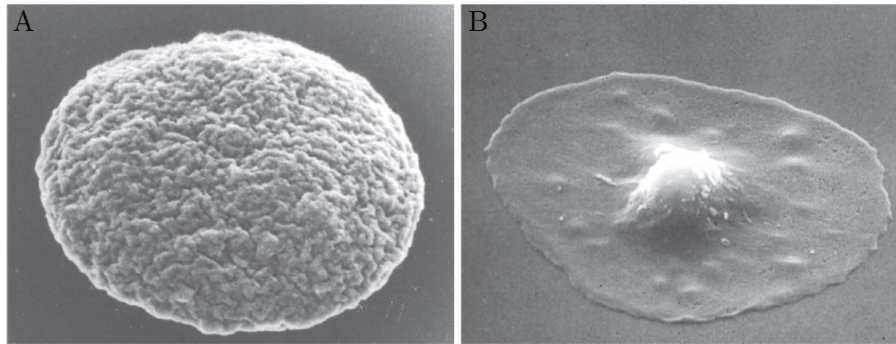


Figure 1.2: During activation, platelets undergo a substantial morphological transformation from (A) a discoid cell shape to (B) a fully spread platelet. Modified with permission from [5] (Copyright 2019, Elviesier).

Circulating platelets are translocated to the site of injury and adhere to subendothelial structures. Platelets undergo structural changes and release procoagulant mediators upon activation, leading to platelet aggregation and thrombus formation (Figure 1.3) [1, 2].



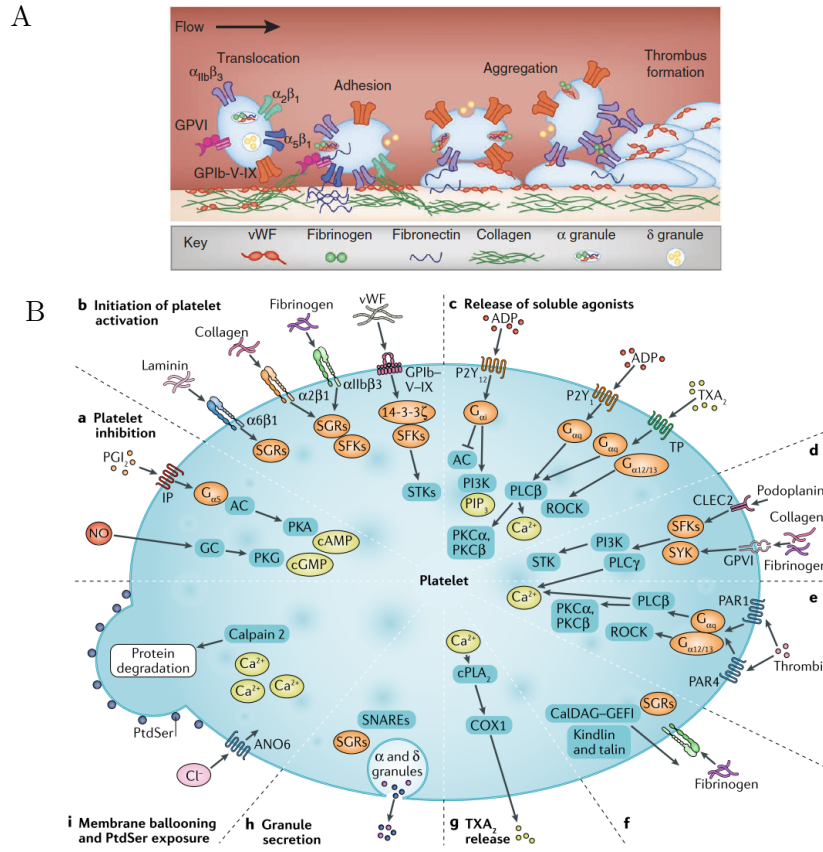


Figure 1.3: The role of platelets in haemostasis: Platelet adhesion, activation, and aggregation cause thrombus formation (A). Reprinted with permission from [2] (Copyright 2011, Nature Publishing Group). Platelet activation as well as preventing activation is under control of multiple signaling events. Reprinted with permission from [1] (Copyright 2018, Springer Nature Limited).

In case of an injured vessel wall, matrix proteins from subendothelium are exhibited to the blood flow. Platelets adhere to the site of injury by several receptor-ligand interactions. The glycoprotein (GP) complex GPIb-V-IX, in particular GPIb $\alpha$ , binds to exposed vWF with rapid on- and off-rates. By this interaction, platelets are recruited to the site of injury by slowing down the cells (platelet tethering), enabling immobilization based on more stable receptor-ligand interactions: GPVI receptor and integrin  $\alpha_2\beta_1$  bind to collagen and integrins  $\alpha_5\beta_1$  and  $\alpha_6\beta_1$  to fibronectin and laminin, respectively. Further platelets are recruited and immobilized onto a growing thrombus based on platelet-platelet interactions involving GPIb-vWF binding as well as binding of integrin  $\alpha_{IIb}\beta_3$  to vWF, fibrinogen, fibrin, and fibronectin (Figure 1.3) [2].

After adhesion, platelets undergo an irreversible morphological and biochemical transformation. Collagen- and fibrinogen-binding to GPVI leads to the activation of phospholipase C $\gamma$  (PLC $\gamma$ ). By the hydrolysis of phosphatidylinositol-4,5-bisphosphate, inositol-1,4,5-triphosphate (IP $_3$ ) is released which initiates the transfer of Ca $^{2+}$  from the DTS to the cy-

toplasma by binding of  $IP_3$  to the receptor  $IP_3R$ , a calcium-selective channel, on the DTS membrane. In case of low  $Ca^{2+}$  levels in DTS, transmembrane protein stromal interaction molecule 1 (STIM1) is translocated to the plasma membrane and increases cytosolic  $Ca^{2+}$  concentration by interaction with the ORAI1 calcium channel [1]. Intracellular  $Ca^{2+}$  levels are also increased by thromboxane ( $TXA_2$ , platelet receptor: thromboxane receptor (TP)) and ADP (platelet receptor:  $P2Y_1$  and  $P2Y_{12}$ ; Figure 1.3 B(c)), as well as thrombin (platelet receptor: PAR1 and PAR4 (protease-activated receptor); Figure 1.3 B(e)) via their G protein-coupled receptors, activation of  $PLC\beta$ , or in case of ADP- $P2Y_{12}$  interaction the activation of phosphoinositide-3-kinases (PI3Ks) as well as inhibition of adenylate cyclase (AC). High cytoplasmatic  $Ca^{2+}$  levels activate cytosolic  $PLA_2$  (c $PLA_2$ ; Figure 1.3 B(g)), mediate, in combination with protein kinase C (PKC), the integrin  $\alpha_{IIb}\beta_3$  activation, the secretion of  $\alpha$ - and dense granules (Figure 1.3 B(h)), and cause morphologic changes by cytoskeletal rearrangement and PS exposure on the outer plasma membrane leaflet [1, 18].

Platelets by themselves contribute to a procoagulant environment through mediator production, release, and thrombin generation. The exposure of PS on the plasma membrane enables the enzymatic cleavage of prothrombin into its active form thrombin by the prothrombinase complex. The secretion of  $\alpha$ - and dense granules releases among others ADP, fibrinogen, and vWF, leading to recruitment and activation of other platelets. In addition, the granule fusion with the plasma membrane causes the increased membrane expression of integrin  $\alpha_{IIb}\beta_3$ . Moreover, thromboxane  $TXA_2$  is produced through the activation of c $PLA_2$ . The c $PLA_2$  cleaves polyunsaturated phospholipids at *sn2* position. Mediated by cyclooxygenase 1 (COX1), released arachidonic acid (AA) is converted into prostaglandin  $G_2$  (PGG<sub>2</sub>) and further into PGH<sub>2</sub>, followed by  $TXA_2$  synthesis by thromboxane synthase [1].

Moreover, platelet activation and aggregation is under regulation of several membrane-located chemokine receptors (CXCR) and their respective ligands (CXCL). Chemokines are released by platelets from  $\alpha$ -granules upon platelet activation. After chemokine-receptor ligation, various downstream signaling pathways are activated, regulating the cellular response in dependence on the involved chemokine receptors and ligands. As an example, the ligation of CXCL12 with CXCR4 causes platelet activation, aggregation, and death, while platelet survival is the result of binding to CXCR7 [22, 23].

Undesired platelet activation and aggregation is prevented by the vascular endothelium through degradation of ATP and ADP by ectonucleotidases, thrombin inactivation by thrombomodulin, and  $PGI_2$  (=prostacyclin) and nitric oxide (NO) release [1]. Endothelial cells produce NO and  $PGI_2$  by oxidation of arginine through nitric oxide synthase and oxidation of AA through COX2 and prostacyclin synthase, respectively. NO binds to the soluble guanylyl cyclase (sGC) after diffusion into the platelet and causes the increase of cyclic guanosine monophosphate (cGMP).  $PGI_2$  increases the level of cyclic adenosine monophosphate (cAMP) by the binding of the prostacyclin receptor (IP) and activation of adenylate cyclase. Both cGMP and cAMP lower the cytoplasmatic  $Ca^{2+}$  concentration and  $IP_3$  synthesis with inhibitory effects on platelet activation.

### 1.1.3 Cardiovascular disease

Haemostasis is a vital mechanism to prevent bleeding as a consequence of vascular injury. Lesions are occluded by a haemostatic plug growing extravascularly (Figure 1.4). However, undesired thrombus formation into the intravascular space causes life-threatening events by reducing or inhibiting the blood flow, e.g. myocardial infarction or stroke, when coronary or cerebral blood circulation is affected [2].

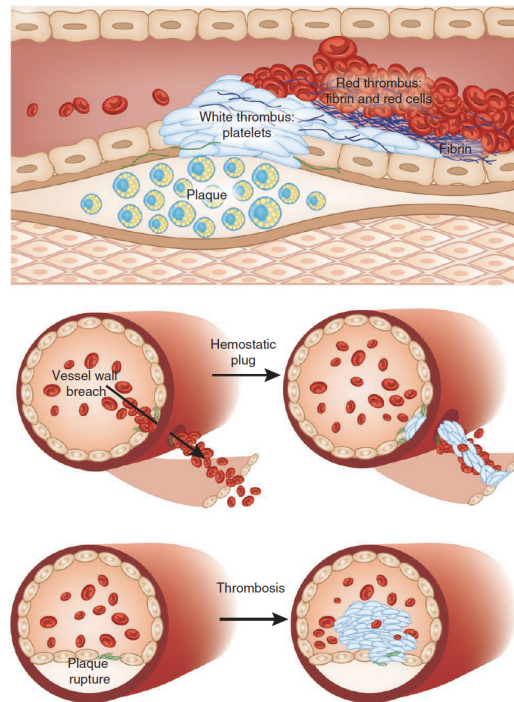


Figure 1.4: The role of platelets in thrombosis: intravascular thrombus formation at a ruptured atherosclerotic plug in contrary to an extravascularly growing thrombus caused by a vascular injury. Plaque rupture induces the formation of a white, platelet rich thrombus followed by the development of a red thrombus based on red blood cells and fibrin. Reprinted with permission from [2] (Copyright 2011, Nature Publishing Group).

Diseases related to the cardiovascular system are responsible for more than 25% of worldwide deaths [24]. In addition, antiplatelet therapy concepts do not prevent patients from further thrombotic events in a reasonable manner since 75% of medicated patients still suffer life-threatening events [25]. Due to the still rising prevalence, research efforts on physiological and pathophysiological mechanisms as well as diagnostic and therapeutic strategies are highly required to reduce atherothrombotic related morbidity and mortality.

Atherothrombosis includes atherosclerosis and arterial thrombosis as two individual diseases being different in pathogenesis and therapy. While atherosclerosis is considered as an inflammatory disease involving endothelial cells, smooth muscle cells, leukocytes, and the adaptive immune system [26], thrombosis is induced by platelets and fibrin [27]. Since dysregulated ad-

hesion of platelets to leukocytes relates thrombotic and inflammatory processes, both diseases cohere in the progression of atherothrombosis [3].

Arterial thrombosis can be considered as an acute event on the basis of a chronic vascular inflammation, i.e. atherosclerosis [2]. Atherosclerotic lesions often develop slowly, thus compensating mechanisms, i.e. collateral vessels, evolve and enable the maintenance of sufficient blood supply. In most of those cases, plaque rupture induced thrombi are nonocclusive with asymptomatic resolution without ischemic lesions of related tissues [28–31].

Through their proinflammatory and thrombus formation properties, platelets contribute to the initiation and propagation of atherosclerotic lesions (Figure 1.5).

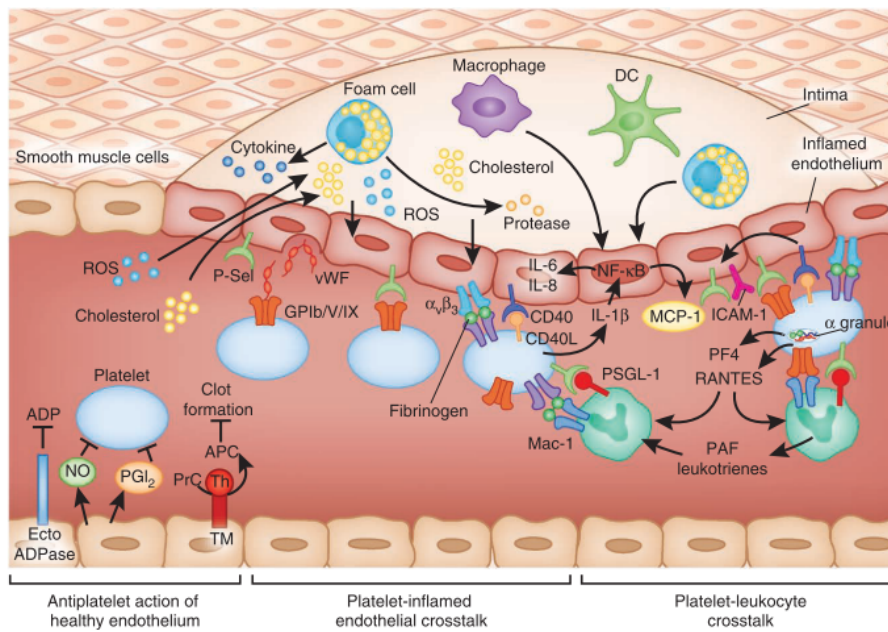


Figure 1.5: Platelet induced progression of atherosclerotic plaques by interactions of platelets with endothelial cells and leukocytes. Physiological antiplatelet actions are inhibited on inflamed endothelium, which expresses adhesive ligands promoting intercellular interactions of endothelial cells and platelets. Activated platelets further recruit leukocytes. Several mediators, released and expressed as a result of intercellular crosstalk, contribute to inflammation progression. Reprinted with permission from [2] (Copyright 2011, Nature Publishing Group).

Under physiologic conditions, the vascular endothelium prevents platelet adhesion by release of NO and PGI<sub>2</sub> as well as expression of ectonucleotidases and thrombomodulin [2]. However, under hyperlipidemic conditions, the altered lipid metabolism leads to endothelial inflammation induced by reactive oxygen species (ROS; superoxide radical anion (O<sub>2</sub><sup>•-</sup>), hydroxyl radical (HO<sup>•</sup>), hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>)) and proatherogenic lipoproteins like oxidized low-density lipoproteins (oxLDL) in the intima and the subsequent recruitment of monocytes, further differentiating into macrophages [32]. Mediated by enhanced uptake of LDL and oxLDL into smooth muscle cells and macrophages, induced by oxLDL-scavenger receptor in-

teractions [33, 34], foam cells accumulate in the intima promoting the intraluminal expression of adhesion ligands. Platelets adhere to the activated endothelium by binding vWF and P-selectin (P-Sel) through GPIb/V/IX. More stable interactions are then promoted by integrin  $\alpha_{IIb}\beta_3$ -fibrinogen-integrin  $\alpha_v\beta_3$  binding, or intercellular adhesion molecule-1 (ICAM-1) [2]. Upon platelet activation, cytokines (interleukin  $1\beta$  (IL- $1\beta$ )) and chemokines (platelet factor 4 (PF4), chemokine ligand 5 (CCL5=RANTES)) as well as proinflammatory mediators (platelet activating factor (PAF) and leukotrienes) are released. Moreover, crosstalk between platelets and inflamed endothelium as well as platelets and leukocytes is enhanced through CD40L (binds to CD40 on endothelial cells) and P-Sel (binds to P-selectin glycoprotein ligand-1 (PSGL-1) on leukocytes) expression on the platelet surface [2, 3, 34]. All those mediators and interactions contribute to inflammation propagation and to an accelerated progression of atherosclerotic lesions.

Consequently, the development of atherosclerosis and arterial thrombosis is correlated and platelets and platelet derived mediators are essential players in both diseases. Therefore, the risk to develop a life-threatening, acute coronary syndrome is even more enhanced in patients with hyperreactive platelets [35–39]. Hyperreactivity is observed in patients suffering e.g. diabetes [40, 41], hypertension [42], or hyperlipidemia [43], all of them cardiovascular risk factors.

## 1.2 Lipids

Lipids are biomolecules with a huge diversity in chemical structure and biological function. Based on the definition from LIPID MAPS (=Lipid Metabolites and Pathways Strategy), lipids are "hydrophobic or amphipathic small molecules that may originate entirely or in part by carbanion based condensations of ketoacyl thioesters and/or by carbocation based condensations of isoprene units" [44].

### 1.2.1 Lipid classification

Since 2003, the LIPID MAPS consortium aims to harmonize and promote lipid research by providing lipid nomenclature rules, databases, tools, and more to support new and experienced researchers. Based on their definition, lipids were classified into the eight categories fatty acyls (FA), glycerolipids (GL), glycerophospholipids (GP), sphingolipids (SP), sterol lipids (ST), prenol lipids (PR), saccharolipids (SL), and polyketides (PK) [44]. Each category includes several classes and subclasses to which individual lipid species are assigned to. All lipid categories are based on one of the two building blocks, either ketoacyl (acetyl or propionyl group) or isoprene groups (Figure 1.6).

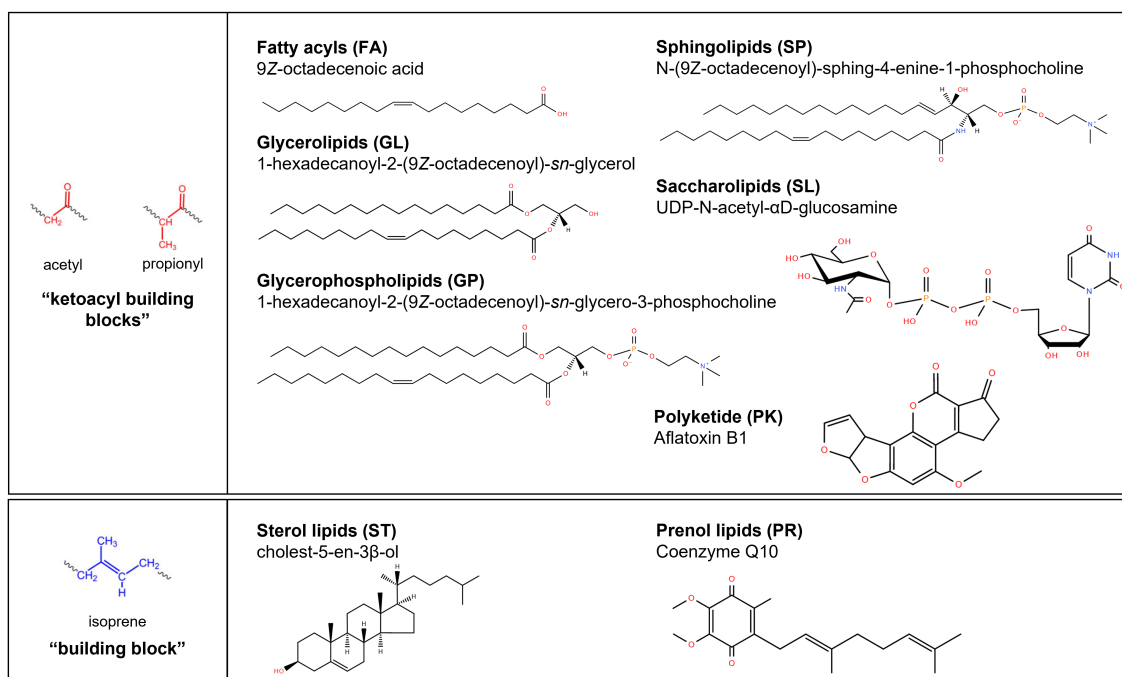


Figure 1.6: Lipid categories and their respective building blocks based on LIPID MAPS lipid definition [44]. Parts were modified and reprinted with permission from [44] (Copyright 2011, Elsevier B.V.).

However, due to the structural diversity, lipid nomenclature is complex and several parameters have to be considered. For simplification, short annotation rules for lipids were proposed [45–47]. The short annotation consists of a letter code indicating the lipid class or subclass and information about the number of carbon atoms and double bonds [46, 47]. As shown in Table 1.1 and corresponding Figure 1.7, depending on the level of structural resolution, rules for annotation of bond types, fatty acyl components, their position, position and configuration of double bonds, and stereochemistry are provided [46, 48]. Consequently, this notation system enables the characterization of lipids as precise as possible in consideration of bioanalytical potentials for structural resolution.

Table 1.1: Levels of lipid notation with increasing degree of structural information based on Koelmel et al. [48] and Liebisch et al. [46] exemplarily explained on 1-hexadecanoyl-2-(9*Z*-octadecenoyl)-*sn*-glycero-3-phosphocholine and N-(9*Z*-octadecenoyl)-sphing-4-enine-1-phosphocholine

| level | structural resolution                     | GP and GL                     | lipid notation<br>SP                                 |
|-------|-------------------------------------------|-------------------------------|------------------------------------------------------|
| 1     | total number of carbons and double bonds  | PC(34:1)                      | SM(36:2;O2)                                          |
| 2     | variable components identified            | PC(16:0_18:1)                 | SM(d18:1/18:1)                                       |
| 3     | positional isomer                         | PC(16:0/18:1)                 | SM(d18:1/18:1)                                       |
| 4     | position of double bond                   | PC(16:0/18:1(9))              | SM(d18:1(4)/18:1(9))                                 |
| 5     | position and configuration of double bond | PC(16:0/18:1(9 <i>Z</i> ))    | SM(d18:1(4 <i>E</i> )(1OH,3OH)/18:1(9 <i>Z</i> ))    |
| 6     | stereochemistry                           | PC(16:0/18:1(9 <i>Z</i> )[R]) | SM(d18:1(4 <i>E</i> )(1OH,3OH)/18:1(9 <i>Z</i> )[R]) |

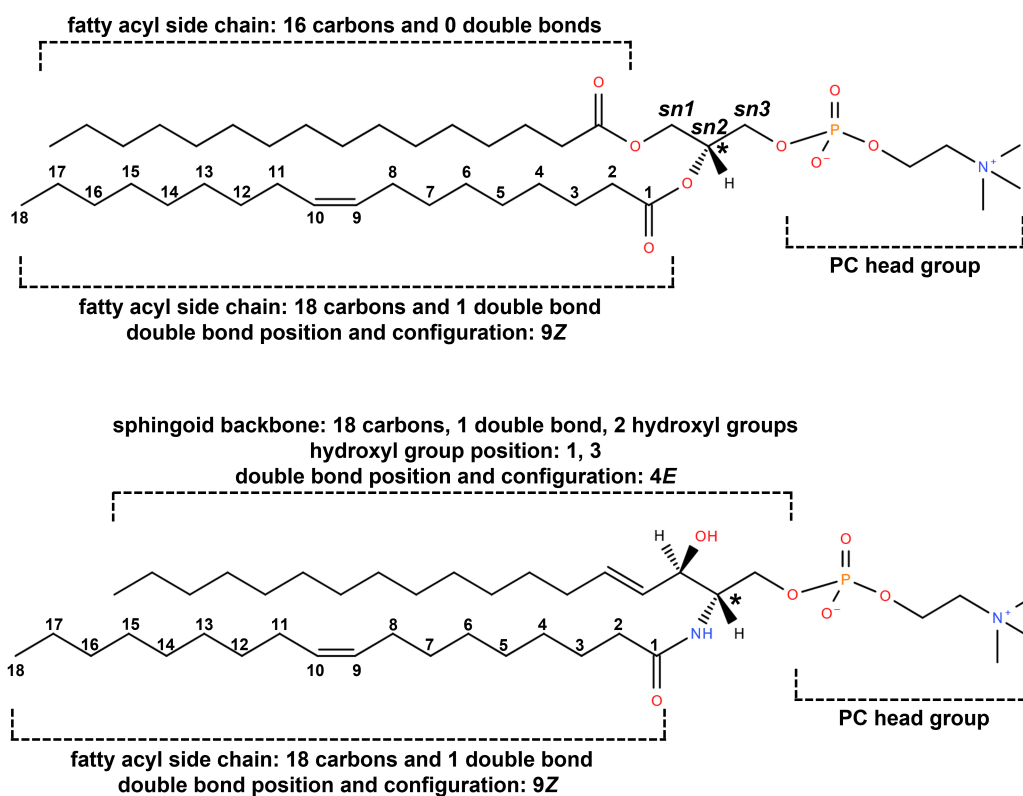


Figure 1.7: Structure of 1-hexadecanoyl-2-(9*Z*-octadecenoyl)-*sn*-glycero-3-phosphocholine (top) and N-(9*Z*-octadecenoyl)-sphing-4-enine-1-phosphocholine (bottom); \* indicates a stereogenic center.

As shown in Table 1.1, the lowest degree of structural information is provided by level 1 only indicating the lipid class with total carbon and double bond number [48]. For glycerolipids and glycerophospholipids, the bond type of fatty acyls to the glycerol backbone is specified for all other than esters, i.e. O-alkyl bond (= "O", e.g. PC(O-34:1)) and O-alk-1-enyl bond (= "P", e.g. PC(P-34:1)). In case of sphingolipids, the number of hydroxyl groups at the sphingoid backbone is indicated with "m", "d", or "t" for mono-, di- or trihydroxylation, respectively [46]. In the next levels, information about the fatty acyl side chains is available with unknown *sn*-position indicated by underscore "\_" (levels 2) and known position with a slash "/" (level 3) between the fatty acyl notations in GL and GP. For the SP notation, the number of carbons and double bonds of sphingoid base and fatty acyl chain are differentiated with a slash "/" [46, 48]. With further information about position of hydroxyl groups, position of double bonds (level 4) combined with their *cis/trans*-geometry (level 5) and configuration on stereochemical center (level 6), lipid structure is resolved completely [48].

### 1.2.2 Biology of lipids

Lipids are vital compounds with manifold biological functions in organisms, especially in cellular structure, energy storage, or as signaling metabolites [49]. Biosynthetically, lipids are either build from ketoacyl groups by Claisen condensation with a carbanion intermediate catalyzed by fatty acid synthase (FAS) or from isoprene units in a carbocation-mediated chain extension (Figure 1.8) [44].



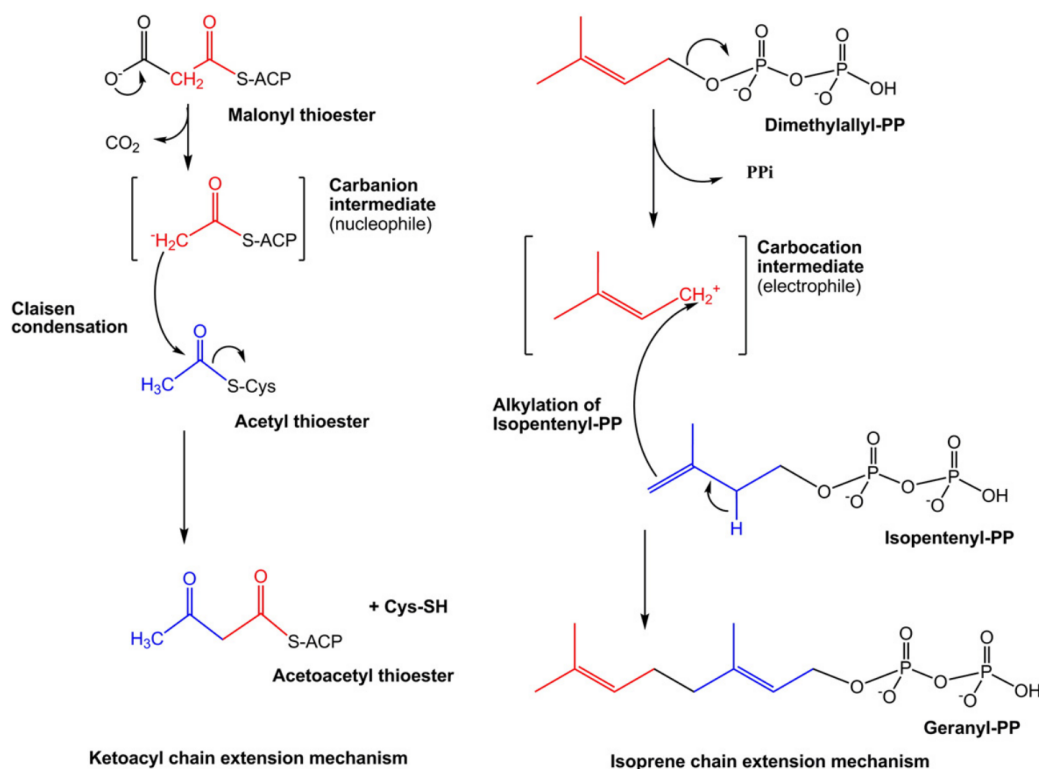


Figure 1.8: Lipid biosynthesis is based on Claisen condensation of ketoacyl moieties or carbocation-mediated chain elongation of the isoprene units dimethylallyl-PP and isopentenyl-PP; ACP = acyl carrier protein. Reprinted with permission from [44] (Copyright 2011, Elsevier B.V.).

In correspondence with their chemical structure, lipids have different functions in cells. Due to their amphipathic properties, glycerophospholipids are membrane-building constituents and the main lipid class found in cellular membranes. Sphingomyelins and cholesterol, which are sphingolipids and sterols, respectively, are also prominent lipids found in membranes [10].

Furthermore, phospholipids are essential players in lipid signaling since they provide polyunsaturated fatty acids (PUFA) released by cytosolic  $\text{PLA}_2$ . Through lipid oxidation, oxylipins are generated from PUFAs, e.g. eicosanoids, further acting as auto- and paracrine pro- or antithrombotic mediators. Enzymatic lipid oxidation is performed initially by the oxygenases lipoxygenase (LOX), cyclooxygenase (COX), or cytochrome P450 (CYP450) [50]. Additional enzymes might be involved downstream, e.g. thromboxane synthase in the production of procoagulant lipid mediator  $\text{TXA}_2$  [51]. However, non-enzymatic oxidation also occurs under oxidative stress through ROS [52]. Other lipids have also signaling functions, e.g. diacylglycerols (DGs), steroids, or acylcarnitines [53].

Another major cellular function of lipids is providing and storage of energy. Basically, energy is obtained by a three step process in mitochondria. From fatty acids, acetyl-CoAs are produced by  $\beta$ -oxidation and transferred to the citric acid cycle. Through electron transfer

to NADH and FADH<sub>2</sub> and their transfer to the respiratory chain in mitochondria, energy is provided as ATP. Besides mitochondria, peroxisomes are involved in  $\beta$ -oxidation since they are specified on oxidation of very long chain fatty acids (VLCFA,  $\geq 22$  carbons) and branched-chain fatty acids [54]. Short, medium, and long chain fatty acids (SCFA,  $\leq 5$  carbons; MCFA, 6 to 12 carbons; LCFA, 13 to 21 carbons) are metabolized in the mitochondria. Fatty acids are stored by ester bonding to glycerol as glycerolipids, e.g. triacylglycerols (TG), in intracellular lipid droplets.

### 1.3 Lipidomics - Techniques and analytical strategies

As presented above, lipids play an essential role in cell integrity and are involved in manifold cellular processes like energy storage and signaling. Thus, dysregulation in the lipid metabolism is related to several diseases, amongst others cardiovascular disease [55, 56]. Consequently, lipid research on physiological and pathophysiological mechanisms as well as diagnostic and therapeutic targets is highly required in preventing, diagnosing, and curating human diseases.

In 2003, lipidomics was introduced as a part of metabolomics [57] and is one of the omics disciplines being highly correlated with the phenotype. Among genomics, transcriptomics, and proteomics, lipidomics is the most recent research field [58]. The term lipidome refers to the total entity of lipids within a compartment, cell, or organism and lipidomics is defined as its structural and functional characterization [55]. A typical lipidomics workflow consists of the steps, i. sample preparation, ii. data acquisition, iii. data processing, and iv. statistical evaluation.

#### 1.3.1 Sample preparation

Lipid extraction is a critical step in lipidomics with respect to lipid coverage and recovery, reproducibility, and suitability for further measurement procedures. Depending on the solvent, the composition of the extracted lipidome varies in terms of lipid classes and amounts recovered from a matrix [59]. Being the initial step in complex lipidomics procedures, improper method selection compromises the entire lipid analysis due to propagation of systematic or random errors.

Most commonly, biphasic extraction procedures are used with solvent mixtures creating an aqueous and an organic layer for the separation of lipids from the water-soluble rest of a sample. Starting from chloroform-based extraction protocols from Folch [60] and Bligh and Dyer [61], several biphasic protocols e.g. using mixtures of hexane-isopropanol-water [62], methyl-tert-butyl-ether(MTBE)-methanol-water [63], or butanol-methanol-water [64] were developed to reduce toxicity, improve performance, and optimize sample preparation procedure. However, phase separation in biphasic protocols is a time-intensive handling step bearing a high contamination risk, especially in protocols with the organic phase as the bottom layer. In contrast to this, monophasic solvent mixtures do not require phase separation steps, thus, they

are of special interest in high-throughput large-scale studies due to their simplified extraction procedure [65, 66]. Basically, the addition of organic, water-miscible solvents causes protein precipitation, thus, the lipid containing supernatant can be removed easily after centrifugation. Despite the simplified sample preparation, the monophasic extraction procedure using a mixture of isopropanol (IPA) and H<sub>2</sub>O showed similar or even better performance in terms of lipid coverage, recoveries, and precision for most of the lipid classes, except of the neutral lipids TGs and CEs exhibiting lower recoveries [66, 67]. Therefore, the monophasic procedure is a very good choice for global and precise but time- and cost-effective lipid extraction in large-scale clinical platelet lipidomics.

### 1.3.2 Liquid chromatography

From an analytical point of view, the qualitative and quantitative analysis of the lipidome is challenging due to the physicochemical and numerical diversity as well as the wide physiological concentration ranges of lipids [55]. Among spectroscopic, chromatographic, and mass spectrometric techniques, the hyphenation of liquid chromatography (LC) with mass spectrometry (MS) emerged to one of the key techniques in lipidomics [58].

The majority of lipidome analysis is based on mass spectrometry driven analytical strategies which are either applied in direct infusion (i.e. shotgun), separation, or imaging approaches [68]. The latter strategy enables the spatial localization and abundance of lipids in tissues [69, 70] while lipid extracts from cells, tissues, or fluids are analyzed by shotgun or separation techniques. Shotgun lipidomics has the main advantages of short analysis times and quantitative robustness since the entire lipid extract is subjected to the MS instrument at once with coionization of added internal standards [71]. However, the simultaneous infusion leads to the major drawbacks, which are large matrix effects and ion suppression [72, 73], lowering the detection sensitivity especially for low abundant species, in-source fragmentation [74], as well as signal interferences of isomers and isobars [75].

Consequently, the reduction of sample complexity is the key solution to encounter direct infusions detriments and is achieved by lipid separation prior MS injection, in general by the application of chromatography. While gas chromatography (GC) is often used for fatty acid analysis after lipid derivatization to volatile fatty acid methyl esters (FAME) [76], liquid chromatography (LC) is the most common technique in lipid separation.

In dependence on the mode of liquid chromatography, either lipid species or lipid class separation is achieved. In supercritical fluid chromatography (SFC), lipids are separated due to their class as well.

#### 1.3.2.1 Lipid species separation

The analysis of complex lipid mixtures is most often performed by reversed-phase liquid chromatography (RPLC) [77]. In RPLC systems, a nonpolar stationary based on silica particles with hydrophobic functionalities, e.g. C8, C18, or phenyl groups, is used in combination with

a polar, water and water-miscible organic solvents containing mobile phase. Analytes are separated by RPLC in dependence on their hydrophobicity, thus, carbon chain length and degree of saturation, leading to resolution of lipid species [78]. Through the application of sub-2  $\mu\text{m}$  fully porous particles [79] or sub-3  $\mu\text{m}$  superficially porous particles [80, 81], the chromatographic performance in terms of speed and efficiency was improved. Especially for large-scale studies, the run time reduction in ultra-high-performance liquid chromatography (UHPLC) is a great advantage to maintain peak capacity for global lipid profiling within shorter analysis times [77].

Through the design of the mobile phase composition, lipid separation as well as detection can be improved. Weak mobile phases, usually used as gradient starting conditions, are composed of water with organic solvents like methanol, acetonitrile, or isopropanol. During the gradient run time, the proportion of strong eluting mobile phases which are based on high amounts of isopropanol or tetrahydrofuran with lower amounts of acetonitrile, methanol, or water, increases [77, 78]. Moreover, mobile phases are modified through low concentrated additives ammonium formate or acetate and their respective acids formic acid and acetic acid, contributing to better chromatographic performance and ionization efficiency [82].

According to the mobile phase, the proper selection of the lipid extract reconstitution solvent is essential for chromatographic performance, i.e. peak shape, and solubility of lipids [77]. However, due to the poor solubility of lipids in aqueous solvents as required in the beginning of a RP gradient, organic solvents are used for extract reconstitution with potential detrimental effects on peak shape of early eluting polar lipids [65, 83].

As a benefit of chromatographic separation prior MS lipid detection, retention time (RT) provides an orthogonal parameter for lipid annotation. In lipid species separation by RPLC, relative retention time correlates with the total number of carbons and double bonds enabling RT prediction [84].

A major disadvantage of lipid species separation is the lack of coelution of internal standards (IS) with their respective analytes. In comprehensive untargeted lipidomics, internal standards are often added to lipid extracts with one IS per lipid class [85]. Without coelution of analyte and IS, ion source conditions, thus matrix effects, are not identical affecting the performance of IS-based normalization and quantification strategies [86].

### 1.3.2.2 Lipid class separation

In contrast to lipid separation based on hydrophobicity, separation due to head group polarity results in lipid class resolution. This is provided by normal phase chromatography (NPLC), hydrophilic interaction chromatography (HILIC) [87], and supercritical fluid chromatography (SFC) [88]. All chromatographic techniques have in common the use of a polar stationary phase, being either bare silica or polar modified silica, e.g. with polyvinyl alcohol [77]. However, based on the mobile phase composition, the three techniques have differences in terms of lipid separation performance and suitability for MS hyphenation.

In NPLC, nonpolar mobile phases are used, mainly composed of organic solvents like hexane or heptane. Due to the high nonpolar solvent content, NPLC separations are less suitable for MS with electrospray ionization (ESI) caused by low ionization efficiencies [78, 89].

In contrast to NPLC, the mobile phase for HILIC is mainly based on acetonitrile with a proportion of water of at least 2.5%. The interaction of water molecules with the stationary phase create a stationary-phase-attached water layer which is essential for analyte separation [77]. The coelution of lipids within a lipid class is advantageous due to coelution of analytes with their respective internal standards, but is also a disadvantage due to peak interferences. In addition, nonpolar lipids like CE and TG are not retarded, thus, those neutral lipids elute more or less with the column void volume.

As a third method in lipid class separations, SFC gained attention due to fast and efficient separation performance [88]. In SFC, CO<sub>2</sub> is applied under supercritical conditions which are above 31°C and 74 bar. Since supercritical CO<sub>2</sub> is miscible with polar solvents, methanol is often used as mobile phase modifier to adjust the polarity [90]. In contrast to HILIC, neutral lipids are retained. However, application of SFC requires special instrumentation (work under high pressure and provide CO<sub>2</sub> flow) compared to other LC techniques [91].

### 1.3.3 Mass spectrometry

Mass spectrometry evolved to the gold-standard technique in omics research due to its great performance in detecting and quantifying analytes, even in complex matrices, with high sensitivity and selectivity. As mentioned above, mass spectrometry is often utilized in lipidomics in shotgun or imaging approaches, or in conjunction with (chromatographic) separation techniques [68]. Basically, the ionization of molecules or molecular fragments after fragmentation enables their separation based on the mass-to-charge ratio ( $m/z$ ) and detection of ion abundance of each species [49, 92]. In the following, the three elemental steps in mass spectrometry, ionization, mass analysis, and detection, are explicated.

#### 1.3.3.1 Ionization

Initially of mass spectrometric analysis, the analytes of interest must be present as gas-phase ions. Thus, ions are generated from the liquid flow in an ion source, which is usually named by the underlying ionization principle. The predominantly used technique is the electrospray ionization (ESI) [93], also used in this lipidomics study. In addition, atmospheric pressure chemical ionization (APCI) as an alternative to ESI and matrix-assisted laser desorption/ionization (MALDI) applied in imaging MS approaches, exist [94].

Using an ESI ion source, a spray is generated from the liquid flow, either coming from direct infusion or preceded chromatographic separation, which is the base of ion formation. In Figure 1.9, the process of ion formation by electrospray ionization is shown.

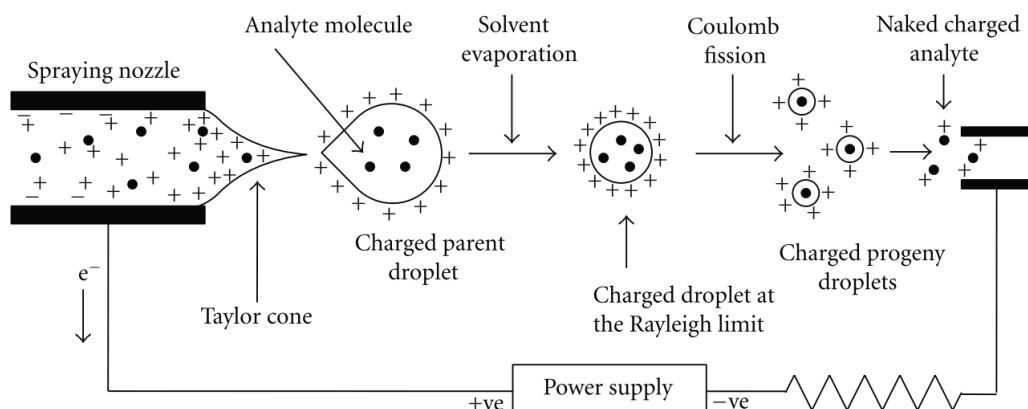


Figure 1.9: Scheme of ion formation by ESI: A needle is used to produce a spray of charged droplets from the liquid flow. Pure analyte ions arise from the droplets through solvent evaporation and Coulombic explosion. Reprinted with permission from [95] (Copyright 2012, Shibdas Banerjee and Shyamalava Mazumdar).

The liquid flow is guided through a needle or capillary being under high voltage, and a Taylor cone is formed at the tip. From this cone, small charged droplets are released, often supported by the application of a nebulizer gas (mostly  $N_2$  or synthetic air). Through temperature induced solvent evaporation, droplet size is decreasing with increasing charge density on its surface. Owing to electrostatic repulsion and when the Rayleigh limit is reached, charged analyte droplets are produced by Coulombic explosion from which single molecular ions are generated after further solvent evaporation [92, 94, 95]. Finally, analyte ions enter the MS instrument through the orifice directed electrostatically. In contrast to the shown axial setup, the ESI needle is usually applied in orthogonal position to the MS entry to prevent contamination by non-ionized molecules [94]. Moreover, a curtain gas flow and high voltage at the curtain plate (plate in front of the orifice; declustering potential) are applied at the orifice protecting the mass analyzer from solvent droplets and removing remaining analyte-solvent clusters, respectively.

ESI is often used in lipidomics analysis since ionization works for almost all lipids [53]. Through electron ejection/capture, (de-)protonation, or adduct formation with charged ions, analyte ions are formed [94]. Usually, single-charged analyte ions like protonated or deprotonated molecular ions  $[M+H]^+$  or  $[M-H]^-$  occur. By adduct formation of lipids with ions from mobile phase additives (e.g.  $[M+NH_4]^+$ ,  $[M+HCOO]^-$ ), even neutral lipids are ionized. Also multiple charged ions occur when analytes are large molecules (e.g. cardiolipins) [96].

ESI provides a reproducible, linear, and sensitive ionization method being sensitive to analyte concentration rather than to the absolute amount [53, 92, 97]. Thus, varying the mobile phase flow rate or the injection volume of the upstream connected chromatographic system enables sensitivity modification.

However, the major drawback of this technique is the significant so-called matrix effect.

When complex matrices like lipid extracts are subjected to the ion source, even after chromatographic separation, analytes compete in droplet formation and ionization [98]. Therefore, coeluting and coionizing analytes affect the signal response commonly leading to ion suppression, but also to ion enhancement [99].

### 1.3.3.2 Mass analyzer

Subsequent to ionization, analyte ions enter the mass analyzer enabling separation, filtering, isolation, or fragmentation according to  $m/z$ -ratio. Various instrument designs based on different principles are available, providing different suitability and performance for qualitative and quantitative analysis.

Mass analyzers are characterized by several parameters, which are mass range, mass accuracy, resolving power, sensitivity, acquisition speed, and linear dynamic range [53, 94, 96]. Upon those, mass accuracy and resolving power are key performance values. Mass accuracy defines the difference between the exact theoretical mass and the observed experimental mass and is typically given in ppm. The resolving power indicates the quality of ion separation with different  $m/z$ -ratios. Thus, it is determined as the  $m/z$ -ratio of an ion related to the mass resolution, i.e. peak width as full width at half maximum (FWHM;  $R=M/\Delta M_{50\%}$ ) [100]. With increasing resolving power, the mass accuracy is enhanced.

Typical mass analyzers are quadrupole, ion trap, Fourier transform ion cyclotron resonance (FT-ICR), orbital ion trap (Orbitrap), or time-of-flight (TOF) instruments, or various combinations of them for tandem MS/MS, either build up from same types of mass analyzers like triple quadrupoles (QqQ) or as hybrid systems like quadrupole-time-of-flight (QTOF). In Table 1.2, performance specifications are given for typical mass analyzers.

Table 1.2: Performance characteristics of typically used mass analyzers [53, 101–103]

| mass analyzer        | resolving power <sup>1</sup> | mass accuracy [ppm] | mass range [ $m/z$ ] | acquisition speed [Hz] | linear dynamic range |
|----------------------|------------------------------|---------------------|----------------------|------------------------|----------------------|
| quadrupole           | 1000                         | 100-1000            | <4000                | 10-20                  | $10^5$               |
| TOF                  | 40000                        | 2-50                | <i>unlimited</i>     | 10-100                 | $10^4$               |
| linear ion trap      | 10000                        | 50-100              | <2000                | 10-30                  | $10^2$ - $10^5$      |
| Orbitrap             | 150000                       | 1-5                 | <6000                | 10-40                  | $10^2$ - $10^5$      |
| FT-ICR               | 1000000                      | 0.1-1               | <10000               | 0.5-2                  | $10^3$ - $10^4$      |
| Sciex TripleTOF 5600 | 35000                        | <3                  | <1250 (Q1)           | 100                    | $10^4$               |

<sup>1</sup> at  $m/z$  1000 and  $m/z$  956 (TripleTOF), respectively.

Based on the resolving power, MS instruments are divided into low resolution (LR) and high resolution (HR) mass spectrometer. LR mass analyzers are often used in quantitative lipidomics, like triple quadrupoles (QqQ) or quadrupole linear ion trap (QTrap) instruments. In contrast, HR-MS, usually applied in qualitative, untargeted analysis, is performed with

instruments having a resolving power above 10000, like QTOF or Orbitraps [100].

In this clinical study, lipidome analysis was performed on the QTOF instrument TripleTOF 5600+ from Sciex equipped with a DuoSpray ion source enabling ESI and APCI. A scheme of the instrumental composition is shown in Figure 1.10.

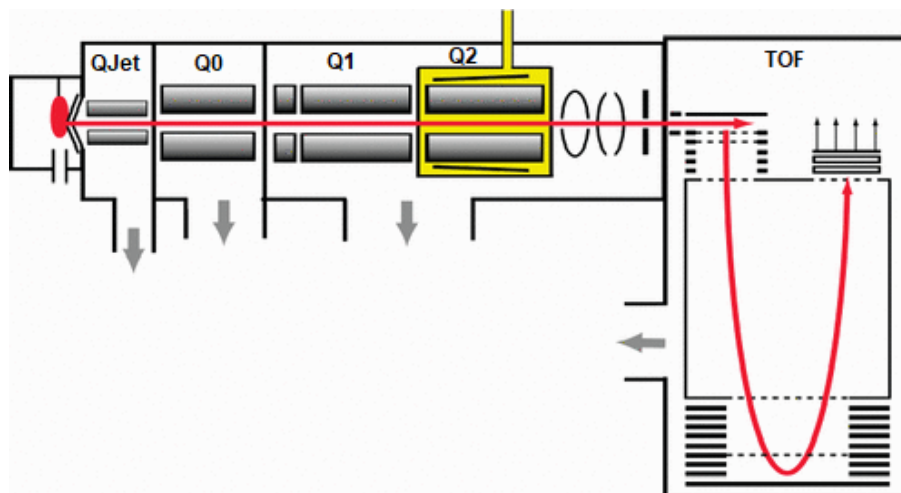


Figure 1.10: Scheme of the Sciex TripleTOF 5600 instrument. The red arrow represents the ion transmission path from the ion source to the detector, grey arrows the vacuum device. Reprinted with permission from [103] (Copyright 2011, American Chemical Society).

The TripleTOF 5600 is a hybrid instrument combining two quadrupoles with a time-of-flight mass analyzer. Further quadrupoles (QJet and Q0) are installed as ion guides.

Through the quadrupoles, ions are transmitted in axial direction in dependence on their  $m/z$ -ratio. Ion transmission and mass filtering is achieved by the application of radio frequency (RF) and direct current (DC) voltage to four hyperbolic rods in parallel orientation with alternating polarities spanning an electromagnetic field (Figure 1.11). In dependence on the combination of RF and DC voltages, only ions of a distinct  $m/z$ -ratio pass the quadrupole on stable, helical trajectories [101, 104], while unstable ion trajectories result in neutralization of ions through touching the rods [105]. Through the increase of RF and DC voltage while maintaining the DC/RF-ratio, a  $m/z$ -range can be scanned resolving ions of different  $m/z$ .



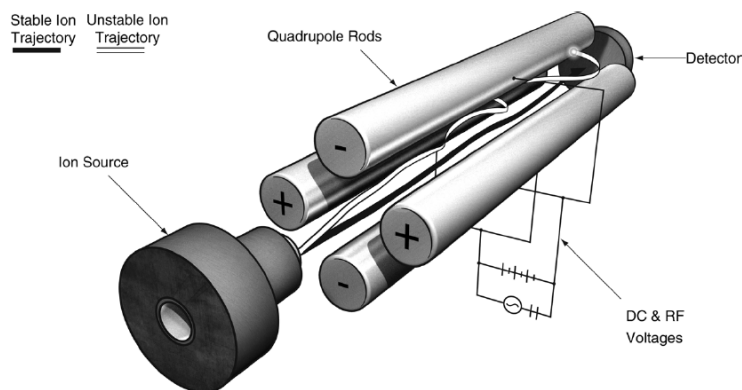


Figure 1.11: Structure of a quadrupole mass analyzer being composed of four rods with applied DC and RF to charge the rods in opposite polarities. Reprinted with permission from [101] (Copyright 2012 Wiley-VCH Verlag GmbH & Co. KGaA).

The relationship of DC and RF magnitudes, resolution, and sensitivity of a quadrupole can be deduced from Mathieu stability diagrams [101] shown in Figure 1.12.

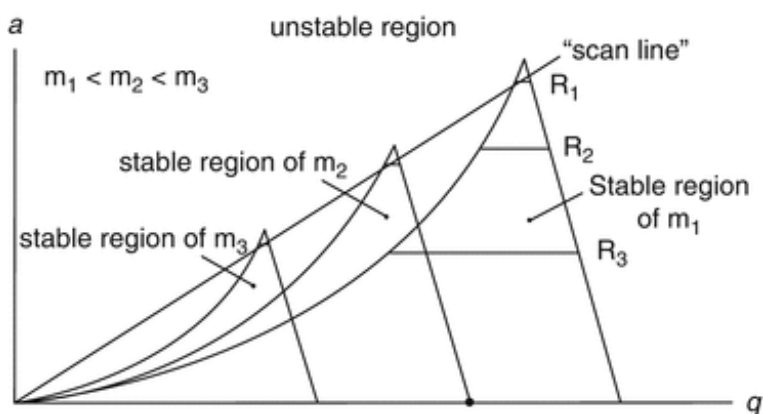


Figure 1.12: Mathieu stability diagram illustrated for three masses with  $m_1 < m_2 < m_3$ . Each  $m/z$  has a stable trajectory for specific DC-RF-combinations with overlaps with other masses. The scan line represents an increase in the magnitude of DC ( $=a$ ) and RF ( $=q$ ) voltages in a stable ratio. In dependence on the intersection of the scan line with the stability areas of the masses, resolution ( $=R_1, R_2, R_3$ ) and sensitivity can be modulated. Reprinted with permission from [106] (Copyright 2017, Springer International Publishing AG).

Ions of distinct masses experience a stable trajectory through the quadrupole in dependence on the magnitudes of DC and RF voltages. These stable areas in the diagram overlap with the stable areas of other masses. The scan line is a result of an increase of the DC and RF voltages magnitude at a constant DC/RF-ratio. Through the variation of the DC/RF-ratio, thus the position and intersection of the scan line with the stable areas, the mass resolution can be modified, i.e. tuning the quadrupole. However, an increase in resolution is compromised with decreased sensitivity [106, 107]. Based on the stability diagram, setting the DC/RF-ratio to

zero (DC=0, "RF-only"), the quadrupole is "open" to transmit all ions [108].

In addition to the first quadrupole, used as a mass filter (Q1), QTOF instruments have a second quadrupole (Q2) working as a collision cell for analyte fragmentation (Figure 1.10). By the collision of traversing ions in the "open" quadrupole with an inert gas like N<sub>2</sub>, ion fragmentation is induced (=collision-induced dissociation (CID)). Since the degree of fragmentation depends on the kinetic energy of the traveling ions, fragmentation patterns can be controlled by the variation of the applied potential, which is called collision energy (CE) [109].

Finally in a QTOF instrument, ions are transmitted to the time-of-flight (TOF) mass analyzer. Basically, ions are separated in dependence on their flight times in a drift tube after acceleration with a distinct potential V. The mass separation due to instrumental parameters and  $m/z$  is well explained by Equation 1.1 with the elementary charge  $e$ , the electric field V, the flight time  $t$ , and the tube length L [53].

$$\frac{m}{z} = 2eV\left(\frac{t}{L}\right)^2 \quad (1.1)$$

Besides the known instrumental settings V and L, a distinct  $m/z$ -ratio depends on the flight time, thus, ions of different masses are separated due to different flight times. Further, ions with lower  $m/z$  travel faster through the drift tube than heavier ions (larger  $m/z$ -ratio) [108]. However, to ensure optimum measurement accuracy and avoid temperature-related variations in tube length or acceleration potential, a temperature controlled environment and regular mass calibration of the TOF mass analyzer is mandatory [110].

As illustrated in Figure 1.10, analyte ions enter the TOF mass analyzer in the acceleration unit. After collecting a distinct portion of ions, they are accelerated together by a defined pulse in orthogonal direction to the TOF entry. Within this region, ions with a longer way to the detector are closer to the pulse source resulting in a stronger acceleration, thus ions with identical  $m/z$  arrive the detector simultaneously. By ion gathering and delayed pulse extraction, loss of resolution due to spatial and kinetic distribution of ions is compensated.

In addition, the ion flight times are elongated by reflecting the accelerated ions with an ion mirror, which is called reflectron (Figure 1.10). The ion mirrors further provide correction of different kinetic energies of ions with identical  $m/z$  since faster ions penetrate deeper prior reflection than ions with lower velocity, thus having a longer way to reach the detector.

### 1.3.3.3 Detectors

Finally, after analyte separation by mass analyzers, a detection system is needed for the conversion of arriving ions into a respective signal, usually, an electric current and its digitization. The former is mainly achieved by electron multipliers, offering detection of incoming ions with signal amplification to obtain sufficient detection sensitivity [111]. Separated ions transmit-

ted from the mass analyzer, the TOF in case of a QTOF instrument, are again accelerated by a high counter potential to hit the detector surface with high kinetic energy. Through this impact, secondary particles are released, i.e. negative ions or electrons in positive mode, positive ions in negative mode [94], which strike the detector surface with further release of electrons. The initiated cascade of impacts results in a signal amplification of  $>10^6$ .

In the TripleTOF 5600+ instrument, the electron multiplier is realized as a microchannel plate (MCP), which is most common in QTOF systems. A MCP is a perforated plate with many small channels providing an increased detector surface for signal amplification. In addition, short electron paths cause a fast detection response [94, 96, 103].

Subsequent to ion detection by the MCP, the resulting electron current has to be converted into a digital signal as data output, either by an analog-to-digital converter (ADC) or a time-to-digital converter (TDC). Due to their good sensitivity [110], the latter are typically used in QTOF instruments. Operating in pulse mode, the TDC is synchronized with the acceleration pulse and registers the arrival time of ions after the last pulse, being transmitted by a discriminator and reaching a distinct threshold [110]. However, TDCs are prone to detector saturation and thus, lower dynamic range. This is caused by dead times during pulse cycles, even if they are short (a few nanoseconds), during which multiple simultaneously arriving ions of the same  $m/z$  are counted as one [110]. To overcome this limitation, multiple TDCs are installed for multichannel detection. The TripleTOF5600+ is equipped with four TDCs [103]. Furthermore, through ion transmission control (ITC), the ion load subjected to the detector is modulated in dependence on the total ion current (TIC) by a lens installed prior the Q0 [94]. A correction factor is finally used to compensate the TDC signal for lower ion load. Due to these techniques, detector linear range is increased.

### 1.3.4 Data acquisition

In dependence on the mode, the mass analyzers are used, data acquisition by mass spectrometry can be divided into targeted and untargeted approaches, basically. Besides the technical application, the approaches differ in the goal of analysis, for which they are applied. Targeted acquisition is used for quantification of known, preselected analytes and its main application is the proof of hypotheses. In contrast, hypotheses are generated through the application of untargeted data acquisition. All detectable analytes, known as well as unknown metabolites, are recorded by this approach, hence, enabling a comprehensive qualitative analysis of a specific biological matrix, e.g. the lipidome of platelets. Although quantification is also possible in untargeted analysis, it is rather relative than absolute like in targeted analysis.

Common to all acquisition modes applied in LC-MS analysis, the proper selection of the total cycle time of a MS measurement is highly important for data quality. The cycle time represents the total required time to acquire one data point including MS and MS/MS experiments as well as instrument times. Thus, for the optimum cycle time selection in LC-MS analysis, detection sensitivity, i.e. low sensitivity with too short cycle times, has to be balanced with

high quality peak representation, i.e. at least ten data points for peak quantification [106, 112]. Usually, in RPLC-MS with C18 UHPLC columns, peak width of around 1 s are obtained, and therefore, the cycle time should not exceed 100 ms.

### 1.3.4.1 Targeted approach

Targeted data acquisition is focused on the analysis of a known set of compounds and is usually performed on low resolution instruments like QqQ or QTrap. Targeted methods, being sufficiently validated, provide the sensitive and selective detection of analytes of interest and enable their absolute quantification. For this purpose, the identity and the fragmentation pattern has to be known in advance. In targeted methods, at least one of the mass analyzers in tandem MS systems is operated in filtering mode for selection of specified  $m/z$  as illustrated in Figure 1.13.

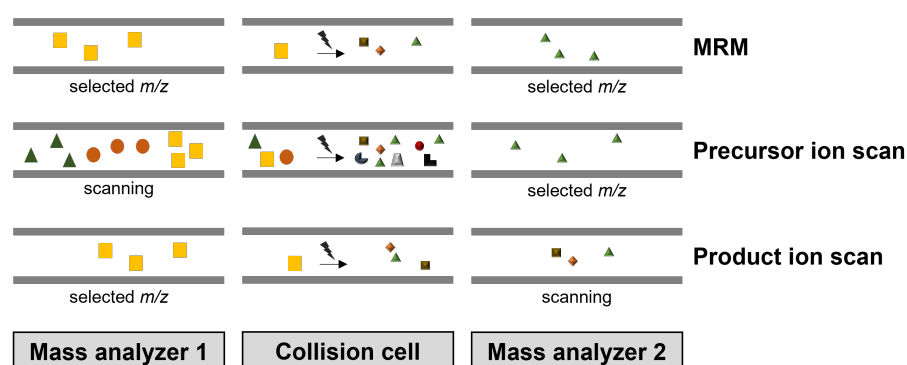


Figure 1.13: Illustration of targeted data acquisition approaches showing multiple reaction monitoring (MRM), precursor ion scan, and product ion scan achieved by  $m/z$  filtering mode of at least one of the two mass analyzers.

Most commonly in metabolite quantification, selected reaction monitoring (SRM), also called multiple reaction monitoring (MRM), is used. In the first mass analyzer, predefined precursor ions of specific  $m/z$  are selected within narrow mass ranges ( $\sim 1$  Da), further fragmented in a collision cell, and selected fragment ions are detected in the second mass analyzer. In case of QqQ, this is all done by quadrupoles. SRM/MRM enables selective and sensitive detection of preselected analytes and their quantification by the use of respective standards at known concentration with short cycle times.

In addition, one of the two mass analyzers can be operated in scanning mode for either a precursor ion scan (i.e. first mass analyzer in scanning mode) or a product ion scan (i.e. second mass analyzer in scanning mode). In precursor ion scans, all precursor ions are detected, exhibiting a specific fragment ion which is selected in the second mass analyzer. This method is suitable for the analysis of a specific lipid class with a characteristic fragment, e.g. PCs and SMs exhibiting  $m/z$  184 [113, 114]. In contrast to this, when the second mass analyzer is operated in scanning mode, all fragments of a selected precursor ion are recorded. The product

ion scan is suitable for the structural analysis of analytes with metabolite identification and quantification.

Targeted methods can be applied for the simultaneous analysis of multiple compounds with a single LC-MS run. However, the number of metabolites is limited due to the maximum cycle time of each MS cycle. To overcome this issue, specific MRMs are acquired only in defined time windows of the chromatogram including the expected retention time of the analyte of interest. Through the scheduled MRM, the number of analytes can be increased by maintaining or increasing the sensitivity of each compound, due to the increase of MRM dwell time [115, 116]. For an additional gain in resolution, especially for a large number of target compounds, high-resolution MRM (MRM<sup>HR</sup>) are performed on high resolution instruments like a QTOF.

The benefit of targeted measurements is the accurate quantification of analyte levels in a given biological matrix. However, although a larger number of targets can be analyzed simultaneously, a global coverage of the lipidome is only achieved by untargeted data acquisition.

#### 1.3.4.2 Untargeted approach with data dependent acquisition

In contrast to the quantitative objective in targeted approaches, untargeted data acquisition is selected for comprehensive lipidome analysis with the aim of hypotheses generation. In principle, data are acquired for all detectable analytes, representing the global lipidome of a biological matrix. However, processing the huge amount of recorded data, especially lipid identification and quantification, is challenging in untargeted research. Basically, untargeted approaches are composed of a survey scan of all detectable precursor ions on MS1 level. Further, analyte fragments are analyzed providing information on MS2 level. For valid lipid identification in untargeted methods, several analytical parameters are required: precursor ion mass and fragmentation pattern, thus MS and MS/MS data, isotope pattern, and, in case of preceded chromatographic separation, retention time information. MS data and MS/MS spectra are compared with reference spectra from databases like LipidBlast [117]. For confident identification, high mass accuracy is required, therefore, untargeted research is performed on high resolution instruments, i.e. QTOF, Orbitrap, FT-ICR [118, 119]. In addition, retention time provides valuable information to support lipid identification, although being more system and method dependent.

Different approaches are available providing untargeted data acquisition on MS and MS/MS level, basically being divided into data dependent acquisition (DDA) and data independent acquisition (DIA).

The principle of data dependent acquisition is a dynamic product ion scan for which the respective precursor ions are selected in a prior survey scan. The selection of precursor ions is information dependent, thus the approach is called information dependent acquisition (IDA) [120, 121]. Based on predefined criteria, which are related to the ion abundance in the survey scan, specific precursor ions are filtered, being further subjected to fragmentation

and acquisition of MS/MS spectra. Usually, 10-20 analyte ions per cycle are selected in IDA when performed on QTOF instruments. The number of triggered precursor ions is adjusted to the instrument acquisition frequency (TripleTOF 5600+: 100 Hz) and the number of datapoints required for sufficient chromatographic peak representation. Data dependent acquisition provides extracted ion chromatograms (EIC) for all analyte ions from the survey scan and high quality spectra on MS2 level of triggered analytes being beneficial for confident lipid identification. However, major drawback of this approach is the lack of global coverage with MS/MS spectra. The precursor selection is not reproducible, since stochastic, therefore, lipid quantification is limited to the MS level without an additional qualifier ion and lipid identification is restricted to the most abundant, triggered analytes. Moreover, less abundant lipids might be missed on MS/MS level.

### 1.3.4.3 Untargeted approach with data independent acquisition

To overcome the limitations from DDA approaches, a comprehensive acquisition of MS and MS/MS spectra of all analytes is required. This is achieved by data independent acquisition (DIA) methods, which are composed of a survey scan followed by fragmentation of all precursor ions for a global analyte coverage on MS and MS/MS level. Regarding the strategy of precursor isolation and MS/MS spectra acquisition, different methods are available.

In an all-ion-fragmentation (AIF, also termed MS<sup>E</sup>) approach [122, 123], two scans of the entire mass range are conducted, first with low and second with high CE to achieve detection of all precursor ions and their respective fragments. Resulting spectra are highly complex, requiring data deconvolution to obtain pure fragment spectra [124].

In contrast to this, MS/MS spectra in MS/MS<sup>ALL</sup> [125] are acquired by ramping the mass extraction window of the first mass analyzer with unit mass resolution over the entire precursor mass range in a product ion scan manner. However, the full comprehensive data acquisition is at the expense of long cycle times restricting the application to shotgun analysis.

Sequential window acquisition of all theoretical fragment ion mass spectra (SWATH) [122, 126] was introduced enabling the application of sequential precursor ion isolation and fragmentation with preceded chromatographic separation. Due to the variable Q1 window size for ion filtering (Figure 1.14), global analyte coverage with MS and MS/MS data is achieved in reasonable cycle times with better selectivity than obtained with AIF approaches. However, the recorded composite MS/MS spectra still require high data processing efforts for data deconvolution and regaining linkage of fragment ions with their respective precursor ion. For this purpose, MS-DIAL software (Mass Spectrometry Data Independent Analysis), specified on processing of DIA data, was developed [117]. In addition, the Q1 window design, i.e. number, width, and schedule of isolation windows, can be adjusted individually to a sample matrix using the swathTUNER software [127]. Based on the chromatographic pattern, Q1 window position and size is optimized for precursor ion density or TIC intensity with respect to a predefined number of so-called SWATH windows.

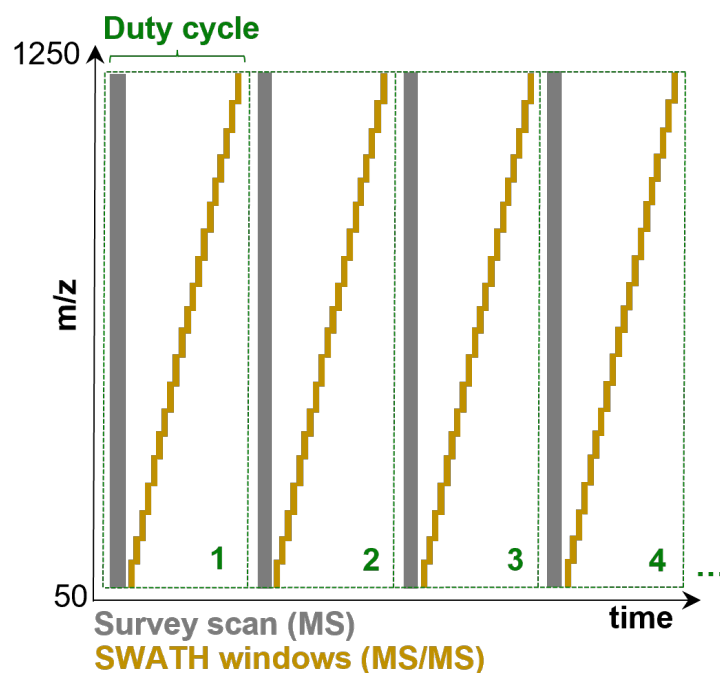


Figure 1.14: Design of DIA experiments using SWATH applying a survey precursor ion scan on MS level and SWATH windows covering the entire mass range for MS/MS data. SWATH windows might be of either fixed or variable size for precursor isolation. The number of SWATH windows as well as the covered mass range has to be adjusted with respect to the total duty cycle time to obtain at least ten datapoints per peak for sufficient peak representation.

SWATH approaches create reliable and reproducible full comprehensive data [128] with global analyte coverage with MS and MS/MS spectra, enabling EIC generation on both levels for all detected analytes. Thus, qualitative and quantitative data analysis is improved in contrast to DDA due to better analyte coverage resulting in higher identification rates and allowing quantification based on the optimal ion in terms of selectivity and sensitivity on either MS or MS/MS level [122, 128, 129]. Moreover, the application of SWATH allows retrospective analysis without the necessity of remeasurement, due to its fully untargeted character. Furthermore, SWATH design can be adjusted to enable untargeted analysis with simultaneous acquisition of specific targets in a targeted manner [83, 115, 130]. However, the simultaneous fragmentation of precursor ions within a defined mass range results in composite spectra, making complex data processing procedures with appropriate software tools mandatory. In addition, recording full comprehensive data creates huge amounts of data [131], with further special demands on bioinformatic solutions and research data management (RDM).

## 1.4 Large-scale lipidomics

Untargeted lipidomics aims to elucidate the lipid profile within a biological matrix of interest enabling hypotheses generation. Since the lipidome is related to the phenotype of an

organism, e.g. humans, lipid profiles are highly individual with a large variety depending on genomic or environmental settings [132]. Thus, for profiling studies in clinical settings with patient samples lacking strictly controlled laboratory conditions, and usually exhibiting smaller biological variation of interest, a large amount of samples is required to accurately and confidently detect biological variation. However, a large-scale clinical study is more than an upscaling of an experiment with limited sample count. Each part, from study design, pre-, peri-, and postanalytical sample and data processing, up to data analysis, requires special considerations and adjustments due to the large amount of samples and usually long time periods, within those studies are conducted. Therefore, the following sections are dedicated to the special requirements in study and experimental design, quality control considerations, and data processing strategies in large-scale omics studies.

#### **1.4.1 Study and experimental design**

For the design of a clinical study and the development of the respective experimental and analytical workflows, the high-throughput of samples has to be considered. Thus, the workflow should be as straightforward and simple as possible, avoiding unnecessary handling steps, ideally being suitable for automatation as well as time and cost effective. Throughout the entire workflow, protocols for sampling, sample handling, sample preparation, and analytical measurement should be standardized in standard operating procedures (SOPs) to maintain similar conditions for data acquisition of all samples of a cohort, contributing to high-quality, robust, and reproducible data [132–134]. Especially in lipidomics, irrespective of an experimental or clinical study design, many efforts are put into the development of guidelines to establish standards for lipid research through the entire workflow from sample collection to data and results reporting. This is the major goal of the Lipidomics Standard Initiative (LSI), in collaboration with LIPID MAPS, offering guidelines and support in the field of lipidomics [44, 135].

Initially of a large-scale study, the sample matrix has to be chosen. In addition to the biological question of investigation, usually requiring a specific sample type, i.e. in the present study being conducted on platelets, complexity of sample collection (e.g. non- or minimal invasive) and sample preprocessing (e.g. serum or plasma, isolation of a specific cell type) as well as storage capacities should be taken into account. Sample handling procedures, i.e. storage, transport, freeze-thaw cycles, sample preparation like lipid extraction, should be properly chosen and standardized to avoid detrimental effects on the sample quality [136]. In detail, sample storage at  $-80^{\circ}\text{C}$  even for several months with one freeze-thaw cycle proved to be suitable [136–138]. To avoid introduction of preanalytical errors and reduce sample variability, beyond sample storage and handling, time of sampling should be carefully chosen and standardized in lipidomics analysis due to the natural fluctuations of the lipidome in dependence on several factors like diet, daytime, medication, or smoking [132].

In lipidomics, the choice of the appropriate lipid extraction method is critical in terms of



lipidome coverage, reproducibility, accuracy, time and cost consumption, but also environmental effects originating from the selection and volume of the extraction solvent [59, 66, 67, 136]. Comparison studies on several commonly applied extraction protocols revealed that a good extraction performance in terms of recovery, precision, and accuracy can be achieved with a simple protocol, both in fluids [67] and cells [66]. The monophasic protein precipitation protocol using isopropanol as extraction solvent performed similar or even better than biphasic extraction protocols like methyl-tert-butyl-ether(MTBE)-methanol-water [63], butanol-methanol-water [64], hexane-isopropanol-water [62], or chloroform-based extraction protocols from Folch [60] and Bligh and Dyer [61] considering simplicity, lipidome coverage, lipid recovery, contamination risk, resource consumption, time efficiency, and, especially in case of the latter mentioned methods, health and environmental hazard. Consequently, monophasic extraction protocols are highly suitable for high-throughput lipid extraction in large-scale applications [66, 67]. Hence, this protocol was selected for the present clinical study on platelet lipidomics. However, further improvement by monophasic method optimization was obtained by Fu et al. [139], enabling faster lipid extraction with good recoveries of polar and even neutral lipids, i.e. TGs and CEs.

Subsequent to sample preparation, lipid extracts are subjected to data acquisition. To handle the large amount of samples, sample preparation and data acquisition are conducted in multiple batches [134]. For this purpose, samples are distributed into batches either by randomized selection or due to the time of sample collection. Within a batch, measurements are conducted on a set of samples from the entire cohort in randomized order without performing instrument maintenance inbetween [132, 133]. Multiple batches are required due to the long time period of sample collection and the impossibility to measure all samples in a single run due to the necessity of regular instrument maintenance and the incidence of column degradation in dependence on injections [133]. Here, it has to be taken into account that data acquisition is often performed in both positive and negative ionization modes for global lipid profiling. Thus, the data are usually collected in two individual runs rather than by polarity switching, reducing the number of samples within an acquisition batch [136]. In this study, the maximum total number of injections within a batch sequence, i.e. clinical samples and system control samples like system suitability test, blank injections, calibration samples, and quality control samples, was 100 for each polarity. Thus, the size of an analytical block is comparable with the batch design of smaller, one-batch studies with similar analytical instrumentation and methods [66, 85, 130]. Instrument cleaning was performed prior each batch. By including system suitability samples, the column performance was assessed to detect column degradation and replace columns in time preventing detrimental effects on data quality of clinical samples. This is a recommended procedure in analytics, however, even more important for large-scale studies [136, 140].

The performance of the applied analytical platform and method should be suitable for high-throughput analysis providing high-quality data, i.e. analytical robustness and reproducibility but also reasonable analysis times per sample [132, 133]. Under appropriate analytical con-

ditions, the analysis time is around 15-20 minutes for UHPLC-MS/MS measurements under RP conditions enabling high sample throughput [117, 136].

In addition, large-scale clinical studies require a controlled data management including meaningful, standardized file naming, secure data storage of large amounts of data, as well as research data management following the FAIR principles (Findability, Accessibility, Interoperability, Reusability) [141, 142].

### 1.4.2 Quality control

In LC-MS based studies being conducted on a large set of samples in several batches over a long period of time, suitable quality assurance is a prerequisite to assess proper analytical performance for confident study results. Through regular quality control, intra- and inter-batch systematic variation can be monitored and, based on respective control samples, batch effect correction is enabled by data processing [134, 136]. Batch effects are a result of intensity and retention time drifts caused by instrumental fluctuations or column aging. Since batch effects are inevitable in large-scale studies, proper correction through the use of quality control samples is mandatory.

The implementation of different types of quality control samples is recommended to account for different types of quality detriments, e.g. system qualification, carry over, signal drifts. Appropriate quality control (QC) samples are system suitability samples, solvent and extraction blanks, standard samples, and quality control samples being representative for the analyzed matrix or a commercially available standard reference material (e.g. NIST SRM 1950) [136, 143]. Quality control samples, mimicking the studied matrix material, are generated in different ways, depending on the study design. Ideally, proportions of sample matrix from each study sample are pooled and processed similarly with the real samples to assess the quality throughout the entire workflow from sample preparation to data processing [134]. However, this is often not feasible, due to the sample matrix (e.g. cells can not be splitted easily, limited sample availability) or data acquisition is started before sample collection is completed. To overcome these issues, QC samples can be prepared as a pool after sample preparation, typically used for cellular sample matrix, or from a representative sample set (e.g. a small proportion from the study cohort) prepared in advance [134, 144]. Matching QC and experimental sample lipid profile is essential to avoid loss of information from the study cohort and maintain or even improve data quality [132, 133].

Quality control samples are implemented into each batch sequence complying different needs. Repetitive QC sample injections are performed at the beginning of a sequence for equilibration purposes (up to ten injections). Further, QC samples are injected regularly (at least every tenth injection [145]) for assessing the intra- and inter-batch reproducibility (i.e. precision  $\leq 30\%$  [132, 136, 146]), junction of multiple batches into one datamatrix, and intra- and inter-batch drift corrections through suitable normalization strategies (as discussed in section 1.4.3) [132, 134, 143].

As a proper tool for quality assessment at a glance, quality control charts were established, being recommended especially in large-scale, long term investigations [147]. In quality control charts, the analytical signal (peak area, intensity, concentration, or retention time) of an analyte is plotted against the injection order. In addition, the average accompanied with warning ( $\pm 2\sigma$ ) and alert ( $\pm 3\sigma$ ) levels are included to visualize whether samples are passing or not passing the acceptance criteria [147]. An exemplified QC chart is depicted in Figure 1.15.

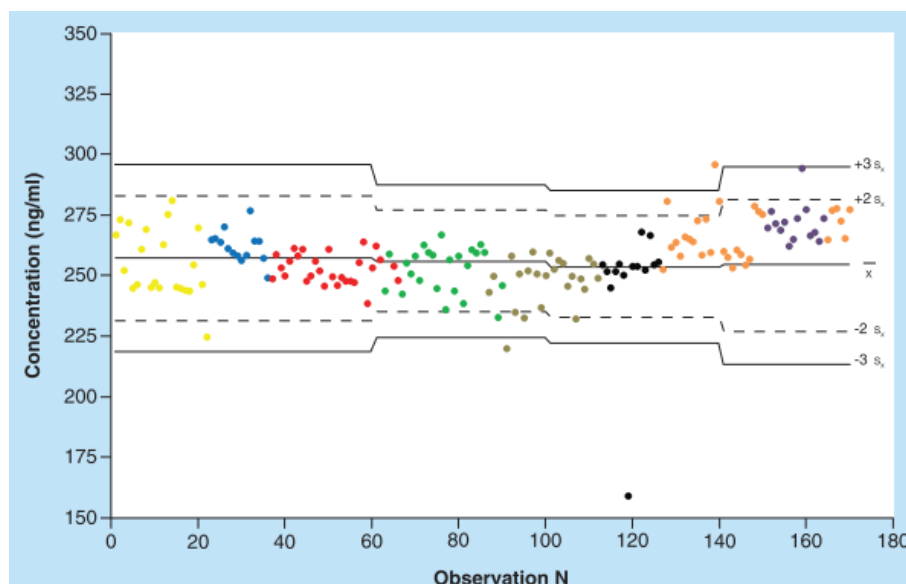


Figure 1.15: Exemplified quality control chart plotting the concentration of analyte X against the injection order. Different colors represent different clinical studies. Reprinted with permission from [147] (Copyright 2017, Future Science Ltd).

### 1.4.3 Data processing strategy

Untargeted acquired lipidomics data are vast and highly complex, thus their handling requires sophisticated bioinformatic strategies and solutions. The objective of data processing is to translate raw chromatographic, mass spectrometric data into metabolite quantities following a complex multi-step processing strategy [53, 148]. As a result, a datamatrix is obtained assigning responses (peak area or peak height) to hundreds and thousands of features (pair of  $m/z$  and retention time) with known or unknown identity, which is further subjected to statistical analysis for conversion into biological knowledge.

The basic data (pre-)processing workflow in lipidomics potentially starts with file conversion from a vendor-specific format into a suitable data type, e.g. an open-format like *ABF* [149] or *mzXML* [150]. The raw data are then subjected to peak picking and peak filtering, including removal of noise by background and blank subtraction [69, 148, 151, 152]. Detected features are aligned between samples within a batch as well as inter-batch-wise in large-scale studies. Further steps are missing values determination and imputation [69, 153–155], isotope

corrections [86, 156], metabolite identification [157], post-acquisitional normalization [69, 158, 159], and quantification [85, 86, 156, 160].

The increasing relevance of lipidomics research is accompanied with a growing demand of suitable bioinformatic tools. LIPID MAPS offers an overview of recommended tools for lipidomics data processing for any acquisition approach, i.e. shotgun or separation techniques with data dependent or data independent acquisition [161]. Adopted from LIPID MAPS, Table 1.3 summarizes suitable open-source tools for lipid identification and quantification in LC-MS/MS analysis. From this table, it becomes evident that there are only a few tools applicable for DIA data processing [131, 161], i.e. MS-DIAL [117] and Lipostar 2.0 [162]. Since its release, MS-DIAL evolved to a key tool in lipidomics processing and is often used as benchmark for newly developed software. Additionally, LipidView provides a vendor-specific software solution for measurements conducted on Sciex instruments [163].

Table 1.3: Tools for processing of untargeted LC-MS/MS lipidomics data providing lipid identification and quantification adopted from LIPID MAPS homepage (<https://www.lipidmaps.org/resources/tools> [18<sup>th</sup> January, 2023]) based on Ni et al. [161]

|                |                                           | Lipid Data<br>Analyzer (LDA) | MS-DIAL            | LipidHunter2 | Lipostar 2.0       | MZmine                               | LPP Tiger |
|----------------|-------------------------------------------|------------------------------|--------------------|--------------|--------------------|--------------------------------------|-----------|
| Identification | MS1                                       | yes                          | yes                | no           | yes                | yes                                  | no        |
|                | MS2 (DDA)                                 | yes                          | yes                | yes          | yes                | yes                                  | yes       |
|                | MS2 (DIA)                                 | no                           | yes<br>(AIF/SWATH) | no           | yes<br>(AIF/SWATH) | no                                   | no        |
|                | Oxidized lipids<br>(coverage)             | yes (4/8)                    | yes (GP)           | no           | yes (8/8)          | no                                   | yes (GP)  |
|                | Spectra matching<br>based ID              | no                           | yes                | no           | yes                | yes                                  | yes       |
|                | Decision rules<br>based ID                | yes                          | yes                | yes          | yes                | yes                                  | yes       |
|                | Spectra<br>annotation                     | yes                          | yes                | yes          | yes                | yes                                  | yes       |
|                | Support multiple<br>adducts               | yes                          | yes                | yes          | yes                | yes                                  | no        |
|                | Correction for<br>in-source fragmentation | no                           | yes                | no           | yes                | no                                   | no        |
|                | Batch processing                          | yes                          | yes                | yes          | yes                | yes                                  | yes       |
| Quantification | Adjustable feature<br>detection           | yes                          | yes                | -            | yes                | yes                                  | -         |
|                | Quantification by<br>Normalization        | Peak volume                  | Peak area          | -            | Peak area          | Peak area                            | -         |
|                | workflows                                 | yes,multiple                 | yes,multiple       | -            | yes,multiple       | yes, after feature-<br>IS-assignment | -         |
|                | Relative<br>quantification                | yes                          | yes                | -            | yes                | yes                                  | -         |
|                | Semi-absolute<br>quantification           | yes                          | yes                | -            | yes                | no                                   | -         |
|                | IS integration                            | yes                          | yes                | -            | yes                | yes                                  | -         |
|                | calibration curve                         | no                           | yes                | -            | yes                | no                                   | -         |

However, besides the suitability of a software tool to handle DIA data, the data processing workflow in large-scale studies has to be applicable for multiple batch acquisitions and a high sample throughput. While latest releases of LC-MS/MS processing software provides batch processing opportunities (see Table 1.3), the simultaneously automated processing of multiple batches with in total several hundreds or thousand samples accompanied with a significant number of quality control samples is critical or impractical in terms of computational capacities, especially due to the large file sizes of DIA data [131]. For example, the main clinical study in this thesis was composed of 22 batches with around 47 clinical and 13 QC samples per batch, thus the entire lipidomics study contained more than 1300 samples. Consequently, the development of a large-scale study data processing workflow is required, implementing and combining the available software tools.

Multi-step data processing [164] of DIA data with MS-DIAL and Lipostar 2.0 offers the full spectrum of data processing steps including background subtraction, peak finding, peak alignment, deisotoping and deconvolution, as well as metabolite identification, quantification, and normalization [117, 162]. Further efforts on data processing solutions for DIA data resulted in the development of LipidMS 3.0 [165], DIAMetAlyzer [166], or DecoMetDIA [167]. LipidMS 3.0 provides an automated lipidomics workflow covering the entire data processing procedure including lipid annotation and quantification, similar to MS-DIAL [165]. The processing strategy of DIAMetAlyzer is based on the library construction using DDA data and applying targeted data extraction of DIA data with automated false-discovery rate control [166]. In contrast, DecoMetDIA is a tool focussing on the deconvolution of highly complex composite spectra, improving identification and structural elucidation of lipid compounds, but lacking a comprehensive processing workflow [167]. However, being the state-of-the-art tool in lipidomics, MS-DIAL was used for the clinical lipidomics study in platelets being presented in this work.

The basis for biological interpretation of LC-MS/MS acquired lipidomics data is the identification of lipids from detected features. Different levels of identification are defined providing increasing confidence of lipid identity from Level 5 to Level 1 with confirmed metabolite structure [168]. The more information is gathered, the more accurate a lipid can be identified and annotated [46, 168, 169]. Using MS data, the accurate mass and the isotopic pattern of a precursor ion give a hint on the sum formula of a feature. Including the fragmentation pattern in MS/MS spectra, spectral matching of experimental data with reference databases, either experimentally acquired or *in-silico* built-up MS/MS libraries, is prerequisite for further structural elucidation. LipidBlast database is an *in-silico* database being released in 2013 and is used as reference database in MS-DIAL based data processing [117, 170]. In addition, orthogonal information like retention time from chromatographic separation or collisional cross section (CCS) in ion mobility measurements is essential to further differentiate compounds with similarities in sum formula and with fragmentation pattern overlaps and increase identification confidence [171]. The final confirmation of a proposed lipid identity is only achieved through a respective reference standard [168].

In all LC-MS/MS assays, response fluctuations caused by random errors occur during sequence acquisition. These unwanted variations, e.g. originating from instrumental (e.g. ionization efficiency of ESI) and environmental (e.g. temperature fluctuations) variations, have to be removed to confidently quantify the detected lipid features [172]. Addressing the introduced fluctuations by suitable normalization strategies is a prerequisite in all LC-MS/MS measurements, however, even more important when analyzing hundreds or thousand samples in multiple batch acquisitions over a long period of time. While normalization in targeted approaches is usually performed by the addition of a respective internal standard, thus correcting the lipid response with the response of the IS, it is more complex and less straightforward in untargeted assays. A plenty of different methods exists, addressing different types of variations, following different normalization strategies, i.e. sample-based scaling methods, IS-based, or QC sample based methods [158, 172, 173]. In scaling approaches, a sample-specific value, e.g. the TIC over all detected [174] or only annotated features [175] or PQN (probabilistic quotient normalization) [176], is calculated and each feature is normalized to this value. Applying spiked IS for response normalization is controversially discussed for untargeted omics, since the number of metabolic features considerably exceeds the number of IS. Thus, IS and feature are lacking chemical similarity and/or coelution causing different matrix effects, or exhibit different response factors due to unequal ionization conditions, e.g. caused by different organic modifier concentrations. However, it is practically not suitable and methodically impossible to have one standard for each lipid feature. Hence, different strategies are evaluated in literature to improve the normalization performance with a limited number of IS [177, 178], basically through optimizing the selection of one or several IS for each feature normalization. Assignment based on chemical properties, e.g. same lipid class [157], is limited to identified features, in contrary to this, assignment based on retention time proximity is universally applicable for all features [179]. Another strategy optimizes the assignment for the minimum coefficient of variation (CV) of normalized features in QCs [178, 180]. Another established method for IS assignment is called B-MIS (best-matched internal standard) applying a minimum relative standard deviation (RSD) in replicates for all possible analyte/IS pairs as the selection criterium [181]. Moreover, more complex, model-based methods are available including NOMIS (normalization using optimal selection of multiple internal standards) [177], CCMN (cross-contribution compensating multiple standard normalization) [182], or RUV (remove unwanted variation) [158]. As the third class of normalization methods, QC-based strategies use the feature responses in regularly injected QC samples to normalize the respective features in real samples [183]. The advantage is that each feature in a sample is normalized by the same feature in a quality control sample, ensuring identical chemical properties and matrix effects. Established QC-based methods are LO(W)ESS (locally (weighted) estimated scatterplot smoothing) [184–186], SVM (support vector machine) [187], or SERRF (systematic error removal using random forest) [159].

The selection of data normalization strategies is a crucial step due to the data manipulation effects. To assist the application of various normalization methods and the performance

evaluation for optimum method selection, several platforms or tools are available [159, 180, 184, 188, 189].



## 2 Objective

The clinical platelet lipidomics studies conducted in this thesis were integrated in a large interdisciplinary research consortium (Collaborative Research Center (CRC)/TR240) dedicated to the investigation of the platelets role in thrombosis, thrombo-inflammation, and immune response in health and disease conditions. In this prospective clinical research, the platelet lipidome was examined under pathophysiological conditions in patients suffering cardiovascular disease. CVD is a highly prevalent disease condition in modern societies, including any blood vessel or heart related diseases, all exhibiting high morbidity and mortality in people. Usually, CVD research and biomarker studies are based on the analysis of plasma metabolites and lipids. However, due to their cellular function, amongst others being involved in thrombotic and thrombo-inflammatory processes, platelets and their lipid metabolites contribute to cardiovascular disease development and progression, thus, the platelet lipidome is of extraordinary interest in disease prevention, diagnosis, and therapy.

The aim of this thesis was the execution, development, and optimization of an analytical workflow for lipid profiling in platelets in large-scale clinical cohorts. Lipidomics analysis was performed by an untargeted RP-UHPLC-ESI-QTOF method in a data independent acquisition using SWATH. The workflow development and optimization was focused on the data processing strategy which is challenging in large-scale multi-batch profiling studies due to large amount of samples and metabolites.

Firstly, a cohort of 120 CAD patients was used for lipidome analysis in multiple batch acquisition dedicated as a prestudy dataset. From an analytical viewpoint, the prestudy cohort should be used to establish a general data processing workflow. Due to computational limitations of simultaneous processing of all samples in an untargeted data extraction procedure, a combined untargeted and targeted data extraction strategy was evaluated. In contrast to a general targeted extraction approach, the goal was to apply unsupervised global lipidome analysis, also covering unknown features and potentially unexpected lipids. For this purpose, an inter-batch feature alignment tool was developed for the preparation of a reference target list for targeted data extraction based on batch-wise untargeted extracted feature lists. The alignment tool was written in VBA being implemented in Excel. Inter-batch retention time adjustment and a suitable alignment strategy under consideration of predefined parameters were required and the proper tool performance had to be evaluated to ensure the correct alignment of identical features from different batches but to avoid misalignments of features not representing the same peak. Further, curation of identified lipids and data normalization strategies had to be included in order to improve the data quality of the resulting lipidomics datamatrix by removal of false-positive lipid annotations and unwanted variance in the analytical response.

Secondly, the developed workflow was optimized and applied to a clinical cohort of 1060 CAD patients being denoted as main study. Untargeted data extraction results in a comprehensive profile of detected features, however, the drawback is a significant proportion of

false-positives. False-positive features originate from the peak picking algorithms in the initial processing of data, being artefacts with detrimental impact on the statistical analysis. Additionally, false-positive lipid annotations occur during automated feature identification based on spectral comparison with reference databases. Thus, the aim was to optimize the workflow regarding qualitative, i.e. data curation of annotated lipids and unknown features, and quantitative data, i.e. missing values handling, isotopic correction, data normalization. Effect evaluation of the processing steps was required to assess the importance of each step to the quality of the final lipidomics datamatrix.

Large-scale clinical lipidomics datamatrices provide a data reservoir for the investigation of several biological and clinical questions. Consequently, both pre- and main study datamatrices were intended to gain a comprehensive view on the platelet lipidome in dependence on disease severity, lipid-lowering statin medication, or for cardiovascular risk assessment with the goal to identify respective lipid biomarkers in platelets, or to investigate the potential of new antithrombotic therapeutic targets.

## 3 Material and methods

### 3.1 Materials

Methanol (MeOH), acetonitrile (ACN), and isopropanol (IPA), all in (Ultra-)LC-MS-grade, as well as formic acid (FA, 98%) were purchased from Carl Roth (Karlsruhe, Germany). Ammonium formate was obtained from Sigma Aldrich (Steinheim, Germany). Ultrapure water was provided by an Elga Purelab purification system (ELGA LabWater, Celle, Germany). Calibration solutions for Sciex TripleTOF 5600+ ESI positive and ESI negative mode were from Sciex (Ontario, Canada). The NIST SRM 1950 plasma was obtained from National Institute of Standards and Technology (NIST, Gaithersburg, MD, USA).

Lipid standards were purchased from either Avanti Polar Lipids (Alabaster, AL, USA) or Cayman Chemicals (Ann Arbor, MI, USA). The pre-mixed standard solutions SPLASH<sup>®</sup> LIPIDOMIX<sup>®</sup> Mass Spec Standard (mixture of 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycero-3-phosphocholine (15:0-18:1(d7) PC), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycero-3-phosphoethanolamine (15:0-18:1(d7) PE), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycero-3-phospho-*L*-serine (15:0-18:1(d7) PS), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycero-3-[phospho-*rac*-(1'-glycerol)] (15:0-18:1(d7) PG), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycero-3-phosphoinositol (15:0-18:1(d7) PI), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycero-3-phosphate (15:0-18:1(d7) PA), 1-(9*Z*-octadecenoyl)(d7)-2-hydroxy-*sn*-glycero-3-phosphocholine (18:1(d7) LPC), 1-(9*Z*-octadecenoyl)(d7)-2-hydroxy-*sn*-glycero-3-phosphoethanolamine (18:1(d7) LPE), cholest-5-en-3 $\beta$ -yl(d7)(9*Z*-octadecenoate) (18:1(d7) Chol Ester), 1-(9*Z*-octadecenoyl)(d7)-*rac*-glycerol (18:1(d7) MG), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycerol (15:0-18:1 (d7) DG), 1,3-dipentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-glycerol (15:0-18:1(d7)-15:0 TG), N-(9*Z*-octadecenoyl)(d9)-*D*-erythro-sphingosylphosphorylcholine (d18:1/18:1(d9) SM), cholest-5-en-3 $\beta$ -ol (Cholesterol(d7)), EquiSPLASH<sup>®</sup> LIPIDOMIX<sup>®</sup> Quantitative Mass Spec Internal Standard (mixture of 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycero-3-phosphocholine (15:0-18:1(d7) PC), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycero-3-phosphoethanolamine (15:0-18:1(d7) PE), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycero-3-phospho-*L*-serine (15:0-18:1(d7) PS), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycero-3-[phospho-*rac*-(1'-glycerol)] (15:0-18:1(d7) PG), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7) *sn*-glycero-3-phosphoinositol (15:0-18:1(d7) PI), 1-(9*Z*-octadecenoyl)(d7)-2-hydroxy-*sn*-glycero-3-phosphocholine (18:1(d7) LPC), 1-(9*Z*-octadecenoyl)(d7)-2-hydroxy-*sn*-glycero-3-phosphoethanolamine (18:1(d7) LPE), cholest-5-en-3 $\beta$ -yl(d7) (9*Z*-octadecenoate) (18:1(d7) Chol Ester), 1-(9*Z*-octadecenoyl)(d7)-*rac*-glycerol (18:1 (d7) MG), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycerol (15:0-18:1(d7) DG), 1,3-dipentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-glycerol (15:0-18:1(d7)-15:0 TG), N-(9*Z*-octadecenoyl)(d9)-*D*-erythro-sphingosylphosphorylcholine (d18:1/18:1(d9) SM), N-pentadecanoyl-*D*-erythro-sphingosine(d7) (C15 Ceramide-(d7)), and LightSPLASH<sup>®</sup> LIPIDOMIX<sup>®</sup> Quantitative Mass Spec Primary Standard (mixture of 1-pentadecanoyl-2-(9*Z*-octadecenoyl)-

*sn*-glycero-3-phosphocholine (15:0-18:1 PC), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)-*sn*-glycero-3-phosphoethanolamine (15:0-18:1 PE), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)-*sn*-glycero-3-phospho-*L*-serine (15:0-18:1 PS), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)-*sn*-glycero-3-[phospho-*rac*-(1'-glycerol)] (15:0-18:1 PG), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)-*sn*-glycero-3-phosphoinositol (15:0-18:1 PI), 1-(9*Z*-octadecenoyl)-2-hydroxy-*sn*-glycero-3-phosphocholine (18:1 LPC), 1-(9*Z*-octadecenoyl)-2-hydroxy-*sn*-glycero-3-phosphoethanolamine (18:1 LPE), cholest-5-en-3 $\beta$ -yl (9*Z*-octadecenoate) (18:1 Chol Ester), 1-(9*Z*-octadecenoyl)-*rac*-glycerol (18:1 MG), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)-*sn*-glycerol (15:0-18:1 DG), 1,3-dipentadecanoyl-2-(9*Z*-octadecenoyl)-glycerol (15:0-18:1-15:0 TG), N-(9*Z*-octadecenoyl)-*D*-erythro-sphingosylphosphorylcholine (d18:1/18:1 SM), N-pentadecanoyl-*D*-erythro-sphingosine (C15 Ceramide) were purchased from Avanti Polar Lipids as well as cholest-5-en-3 $\beta$ -yl(d7) (9*Z*-octadecenoyl) (18:1 (d7) Chol Ester), cholest-5-en-3 $\beta$ -ol(d7) (Cholesterol(d7)), 5*Z*,8*Z*,11*Z*,14*Z*-eicosatetraenoic acid (d11) (AA(d11)), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)(d7)-*sn*-glycero-3-phosphate (15:0-18:1 (d7) PA), sphing-4-enine-1-phosphocholine(d9) (LysoSM d18:1(d9)), cholest-5-en-3 $\beta$ -yl (9*Z*-octadecenoyl) (18:1 Chol Ester), 1-pentadecanoyl-2-(9*Z*-octadecenoyl)-*sn*-glycero-3-phosphate (15:0-18:1 PA), cholest-5-en-3 $\beta$ -ol (Cholesterol), and sphing-4-enine-1-phosphocholine (LysoSM d18:1).

From Cayman Chemicals 5*Z*,8*Z*,11*Z*,14*Z*-eicosatetraenoic acid(d8) (AA(d8)), N-(octadecanoyl)-sphing-4-enine(d7) (d18:1/18:0(d7) Cer), O-hexadecanoyl-*L*-carnitine (d3), sphing-4-enine-1-phosphate d18:1(d7), 5*Z*,8*Z*,11*Z*,14*Z*-eicosatetraenoic acid (AA), O-hexadecanoyl-*L*-carnitine, sphing-4-enine-1-phosphate d18:1 were obtained.

## 3.2 Methods

### 3.2.1 Sample preparation and data acquisition

#### 3.2.1.1 Blood sampling

Blood collection for platelet isolation from cardiology patients suffering coronary artery disease (CAD) was performed at the clinic of the University of Tübingen. The study, following the ethic standards of the Declaration of Helsinki, was approved by the ethics committee of the Medical Faculty of the University of Tübingen (270/2011B01) and patients signed informed consent.

Patients were admitted to hospital due to symptomatic CAD and further diagnosed by coronary angiography within 24 h. Two different sample cohorts were included in this clinical lipid analysis. For the first cohort (= prestudy), blood samples from 139 patients were collected from arterial blood during percutaneous coronary intervention (PCI) prior heparin administration from February to May 2017. Due to the platelet count, only 120 samples were subjected to lipidomics analysis. In the second cohort (= main study), samples were collected postinterventional within 24 h after PCI from venous blood. In total, samples from 1060 patients were collected from March 2018 until March 2021. Finally, 1011 samples were subjected to lipidomics analysis.

### 3.2.1.2 Platelet isolation

Platelet isolation was performed immediately after sampling following the published protocols [130, 190]. Blood was collected in acid-citrate-dextrose (ACD) buffer (60 mL in total, blood:buffer 4:1 (v/v), ACD buffer: 12.5 g sodium citrate, 6.82 g of citric acid, 10 g glucose, 500 mL distilled water, adjusted to pH 4.69 with NaOH). After centrifugation at 330 x g for 10 min (without acceleration and deceleration), obtained platelet-rich plasma was mixed with Tyrodes-HEPES buffer (2.5 mM HEPES, 150 mM NaCl, 1mM KCl, 2.5 mM NaHCO<sub>3</sub>, 0.36 mM NaH<sub>2</sub>PO<sub>4</sub>, 5.5 mM glucose, 1 mg/ml BSA, pH 6.5), centrifuged again at 240 x g for 10 min (without acceleration and deceleration) and supernatant was removed. After careful resuspension of the platelet pellet in 5 mL Tyrodes-HEPES buffer pH 6.5, platelet suspension was centrifuged at 430 x g for 10 min. The supernatant was discarded and the pellet resuspended in 1 mL Tyrodes-HEPES buffer pH6.5 for platelet count determination. The platelet suspension was divided into aliquots with either variable or fixed platelet count ( $3 \times 10^8$  (prestudy cohort),  $2 \times 10^8$  (main study cohort)) platelets per aliquot, depending on the sample. After final centrifugation at 420 x g for 5 min at 4°C, platelet pellets were stored at -80°C until lipid extraction.

### 3.2.1.3 Lipid extraction

Platelet lipid extraction was based on a monophasic precipitation protocol using isopropanol (IPA) as published elsewhere [191]. Platelet samples were divided into batches of maximum 47 samples each by randomized assignment or due to the chronology of sampling. In total, the prestudy samples were distributed into three batches all processed in June/July 2018 and the main study had 22 batches being processed mainly continuously from April 2020 to April 2021.

A volume of 5 mL ice-cold IPA/H<sub>2</sub>O-mixture (9:1, v/v) was added to each sample as well as a mixture of internal lipid standards dissolved in methanol (MeOH). For the prestudy cohort, final lipid standard concentrations were 4% SPLASH<sup>®</sup> LIPIDOMIX<sup>®</sup>, 100 ng/mL AA(d8), and 300 ng/mL d18:1/18:0(d7) Cer per  $3 \times 10^8$  platelets per pellet. In main study samples, the lipid standard mixture contained EquiSPLASH<sup>®</sup> LIPIDOMIX<sup>®</sup> (0.5%), AA(d11) (100 ng/mL), palmitoyl-*L*-carnitine(d3), 15:0-18:1(d7) PA, LysoSM d18:1(d9) (all 500 ng/mL), Cholesterol(d7) (2000 ng/mL), sphingosine-1-phosphate d18:1(d7) (5000 ng/mL), and 18:1(d7) Chol Ester (10000 ng/mL: 500 ng/mL from LIPIDOMIX<sup>®</sup> and 9500 ng/mL spiked internal standard) adjusted for  $2 \times 10^8$  platelets per pellet. Internal standard mixtures were prepared as 10-times concentrated stock solutions, stored at -20°C, and diluted 1:10 with MeOH prior use. A blank extraction and lipid extraction of NIST SRM 1950 plasma (30 µL plus 120 µL extraction solvent with lipid standards) were also included. After vortexing twice for 10 s with ultrasonication for 2 min in between, samples were incubated shaking (500 rpm) on ice for 1 hour with ultrasonication (á 2 min) every 12 min. The supernatant was collected after centrifugation (2740 x g, 10 min, 4°C) and dried over-night under nitrogen in a centrifugal

evaporator (GeneVac EZ-2, SP Industries, Warminster, PA, USA). Dried lipid extracts were reconstituted with MeOH (prestudy:  $3 \times 10^8$  per 100  $\mu\text{L}$  MeOH, main study:  $2 \times 10^8$  per 100  $\mu\text{L}$  MeOH). After another cycle of vortexing (30 s), ultrasonication (2 min), and centrifugation ( $2740 \times g$ , 3 min,  $4^\circ\text{C}$ ), the supernatant was transferred into HPLC vials and subsequently subjected to LC-MS measurement.

#### 3.2.1.4 Preparation of quality control samples

For the preparation of an identical quality control (QC) sample matrix for all batches of the main study, 131 platelet pellets were pooled after reconstitution in 1 mL IPA/H<sub>2</sub>O-mixture (9:1, v/v). After vortexing for 30 s, the pool was distributed into 48 aliquots and another 2 mL extraction solvent were added to each aliquot to obtain a total volume of 5 mL extraction solvent and proceeded with the monophasic lipid extraction protocol as described above.

Due to the limited sample amount in the prestudy cohort, QC samples for all batches of the prestudy were prepared as a pool of lipid extract from the first batch samples.

#### 3.2.1.5 Preparation of calibration samples

Six-point calibration series were included to each batch in the main study. Thus, a calibrant stock solution, which was equivalent with the highest calibrant concentration of the calibration series, was prepared and diluted prior data acquisition of each batch. For the stock solution, unlabeled primary standards (=calibrants) and their correspondant isotopically labeled internal standards (IS) were mixed with concentrations shown in Table 3.1.

Table 3.1: Concentration of calibrants and internal standards in stock solution and six-point calibration series

| calibrants                           | concentrations [% or ng/mL] <sup>2</sup> |        |       |      |      |       |
|--------------------------------------|------------------------------------------|--------|-------|------|------|-------|
|                                      | Cal6(=stock)                             | Cal5   | Cal4  | Cal3 | Cal2 | Cal1  |
| LightSPLASH® LIPIDOMIX® <sup>1</sup> | 30                                       | 10     | 1     | 0.1  | 0.01 | 0.001 |
| AA                                   | 30000                                    | 10000  | 1000  | 100  | 10   | 1     |
| palmitoyl- <i>L</i> -carnitine       | 30000                                    | 10000  | 1000  | 100  | 10   | 1     |
| 15:0-18:1 PA                         | 30000                                    | 10000  | 1000  | 100  | 10   | 1     |
| Cholesterol                          | 250000                                   | 83333  | 8333  | 833  | 83   | 8.3   |
| 18:1 Chol Ester                      | 470000                                   | 156667 | 15667 | 1567 | 157  | 15.7  |
| LysoSM 18:1                          | 30000                                    | 10000  | 1000  | 100  | 10   | 1     |
| sphingosine-1-phosphate              | 30000                                    | 10000  | 1000  | 100  | 10   | 1     |

<sup>1</sup> Concentration given as percentage.

<sup>2</sup> Final concentration of internal standards was identical for all calibrant concentrations: EquiSPLASH® LIPIDOMIX®=0.5%, AA(d11)=100 ng/mL, palmitoyl-*L*-carnitine(d3)=500 ng/mL, 15:0-18:1(d7) PA=500 ng/mL, Cholesterol(d7)=2000 ng/mL, 18:1(d7) Chol Ester=10000 ng/mL, LysoSM 18:1(d9)=500 ng/mL, sphingosine-1-phosphate(d7)=5000 ng/mL.

The calibrants stock solution was stored at  $-20^{\circ}\text{C}$ . Prior acquisition of each batch data, the calibration series was prepared by diluting the stock solution with internal standard mix to the distinct calibrants concentration levels.

### 3.2.1.6 RP-UHPLC-ESI-MS/MS and sequence design

Data acquisition was performed with an Agilent 1290 Infinity Series UHPLC system (Agilent, Waldbronn, Germany) coupled with a Sciex TripleTOF 5600+ series mass spectrometer with a DuoSpray ion source (Sciex, Concord, Ontario, Canada). A CTC Pal HTC-xt autosampler (CTC Analytics, Zwingen, Switzerland) was used for sample application to the LC-system. An Acquity UPLC CSH C18 ( $130\text{\AA}$ ,  $1.7\text{ }\mu\text{m}$ ,  $2.1\text{ mm} \times 100\text{ mm}$ ) column with an Acquity UPLC CSH C18 VanGuard pre-column ( $130\text{\AA}$ ,  $1.7\text{ }\mu\text{m}$ ,  $2.1\text{ mm} \times 5\text{ mm}$ ) (Waters, Eschborn, Germany) were applied for reversed-phase liquid chromatography.

LC-MS conditions were based on a published method from Tsugawa et al. [117]. For lipid separation, a gradient elution was performed using ACN/ $\text{H}_2\text{O}$  (6:4, v/v) with 10 mM ammonium formate and 0.1% formic acid as mobile phase A and IPA/ACN/ $\text{H}_2\text{O}$  (90:9:1, v/v/v) with 10 mM ammonium formate and 0.1% formic acid as mobile phase B. The gradient parameters were as follows: 15% B, 30% B at 2 min, 48% B at 2.5 min, 82% B at 11 min, 99% B at 11.5 min, 99% B at 12 min, 15% B at 12.1 min, 15% B at 15 min for re-equilibration. Samples were kept at  $4^{\circ}\text{C}$  in the autosampler and injected with  $5\text{ }\mu\text{L}$  and  $3\text{ }\mu\text{L}$  in positive and negative ion mode, respectively. The column temperature was  $65^{\circ}\text{C}$  and the flow rate was set to  $0.6\text{ mL/min}$ .

The MS parameters of the TripleTOF are listed in Table 3.2.

Table 3.2: MS parameters

| MS parameter                      | positive mode                   | negative mode                   |
|-----------------------------------|---------------------------------|---------------------------------|
| mass range [ $m/z$ ]              | 50-1250                         | 50-1050                         |
| Ion source gas 1                  | 60 psi                          | 60 psi                          |
| Ion source gas 2                  | 60 psi                          | 60 psi                          |
| Curtain gas (CUR)                 | 35 V                            | 35 V                            |
| Ion spray voltage floating (ISVF) | 5.5 kV                          | -4.5 kV                         |
| Temperature (TEM)                 | $350^{\circ}\text{C}$           | $350^{\circ}\text{C}$           |
| Declustering potential (DP)       | 80 V                            | -80 V                           |
| RF transmission                   | $m/z$ 80: 50%<br>$m/z$ 200: 50% | $m/z$ 80: 50%<br>$m/z$ 200: 50% |

Data acquisition was performed in a data independent acquisition (DIA) mode using sequential window acquisition of all theoretical fragment ion mass spectra (SWATH). Within a period cycle time of 720 ms, a full precursor ion scan (TOF-MS, 50 ms accumulation time) and 20 MS/MS experiments with variable Q1 window size (SWATH windows, each 31 ms accumulation time) were acquired (Table 3.3). The SWATH windows were optimized for the platelet

lipidome based on the chromatographic pattern of a platelet lipid extract and SwathTuner software [127].

Table 3.3: Design of MS (TOF) and MS/MS (SWATH) experiments

| exp. | type  | positive mode |           |                 | negative mode |           |                 |
|------|-------|---------------|-----------|-----------------|---------------|-----------|-----------------|
|      |       | start[m/z]    | stop[m/z] | CE $\pm$ CES[V] | start[m/z]    | stop[m/z] | CE $\pm$ CES[V] |
| 1    | TOF   | 50            | 1250      | 10              | 50            | 1250      | -10             |
| 2    | SWATH | 50            | 214.6     | 45 $\pm$ 15     | 49.5          | 342.2     | -45 $\pm$ 15    |
| 3    | SWATH | 213.6         | 281.8     | 45 $\pm$ 15     | 341.2         | 453.6     | -45 $\pm$ 15    |
| 4    | SWATH | 280.8         | 390.7     | 45 $\pm$ 15     | 452.6         | 480.8     | -45 $\pm$ 15    |
| 5    | SWATH | 389.7         | 480.4     | 45 $\pm$ 15     | 479.8         | 507.8     | -45 $\pm$ 15    |
| 6    | SWATH | 479.4         | 509       | 45 $\pm$ 15     | 506.8         | 532.3     | -45 $\pm$ 15    |
| 7    | SWATH | 508           | 536.5     | 45 $\pm$ 15     | 531.3         | 566.8     | -45 $\pm$ 15    |
| 8    | SWATH | 535.5         | 610.6     | 45 $\pm$ 15     | 565.8         | 617.4     | -45 $\pm$ 15    |
| 9    | SWATH | 609.6         | 677.1     | 45 $\pm$ 15     | 616.4         | 687.1     | -45 $\pm$ 15    |
| 10   | SWATH | 676.1         | 709       | 45 $\pm$ 15     | 686.1         | 715       | -45 $\pm$ 15    |
| 11   | SWATH | 708           | 735.1     | 45 $\pm$ 15     | 714           | 744.1     | -45 $\pm$ 15    |
| 12   | SWATH | 734.1         | 759.1     | 45 $\pm$ 15     | 743.1         | 755.1     | -45 $\pm$ 15    |
| 13   | SWATH | 758.1         | 773.1     | 45 $\pm$ 15     | 754.1         | 776       | -45 $\pm$ 15    |
| 14   | SWATH | 772.1         | 790.2     | 45 $\pm$ 15     | 775           | 794.6     | -45 $\pm$ 15    |
| 15   | SWATH | 789.2         | 811.2     | 45 $\pm$ 15     | 793.6         | 807.6     | -45 $\pm$ 15    |
| 16   | SWATH | 810.2         | 827.2     | 45 $\pm$ 15     | 806.6         | 830.3     | -45 $\pm$ 15    |
| 17   | SWATH | 826.2         | 856.2     | 45 $\pm$ 15     | 829.3         | 840.1     | -45 $\pm$ 15    |
| 18   | SWATH | 855.2         | 884.3     | 45 $\pm$ 15     | 839.1         | 859.2     | -45 $\pm$ 15    |
| 19   | SWATH | 883.3         | 915.9     | 45 $\pm$ 15     | 858.2         | 889.1     | -45 $\pm$ 15    |
| 20   | SWATH | 914.9         | 983.7     | 45 $\pm$ 15     | 888.1         | 924.6     | -45 $\pm$ 15    |
| 21   | SWATH | 982.7         | 1250      | 45 $\pm$ 15     | 923.6         | 1050.5    | -45 $\pm$ 15    |

In all batches, the set and order of system injections, quality control sample injections, and biological sample injections was similar. The general sequence design is shown in Figure 3.1. Each sample block had up to 25 samples acquired in randomized order with QC injections after every fifth sample. The three blocks of 6-point-calibration series were only included in main study data acquisition but not in prestudy batches.



Figure 3.1: The sequence design of a batch: starting block included 2x solvent injections (MeOH), 3x column equilibration samples, 1x blank extraction injection, 3x system suitability test (SST) samples. The two sample blocks included up to 25 samples each and QC injections after every fifth sample. Through the entire sequence, 4 NIST plasma lipid extract injections and in total 12 QCs were distributed.

Included in the starting block, a system suitability test was performed prior biological sample injection to assess proper performance of the LC-MS system. For this purpose, 0.5%



EquiSPLASH<sup>®</sup> LIPIDOMIX<sup>®</sup> diluted in MeOH was injected as triplicates and LC-MS parameters (retention time, peak width<sub>1/2</sub>, peak area, peak height) were compared with reference values. Based on the standard deviation ( $\sigma$ ) of the reference values, warning ( $\pm 2\sigma$ ) and control levels ( $\pm 3\sigma$ ) were calculated for all internal standards, and the SST was passed when 95% of the parameters and internal standards matched the criteria.

### 3.2.1.7 Method validation

The applied method for lipidomic analysis was validated following recommendations for targeted methods of the U.S. Food and Drug Administration (FDA) [146] and suggested procedures from Matuszewski et al. [192]. For validation, linear range, limit of detection (LOD), limit of quantification (LOQ), accuracy, precision, recovery (RE), matrix effect (ME), and process efficiency (PE) were determined for primary standards (calibrants = 19 lipid standards) covering the study-related lipid classes. All validation samples were prepared from a platelet pool obtained from 16 individual platelet pellets. Due to the limited sample matrix, validation samples were prepared as triplicates with 2-times injections of each sample during data acquisition to obtain six repetitive measurements.

Recovery, matrix effect, and process efficiency were determined as ratios from pre- and postspiked lipid extracts and calibrants in neat solution (Equations 3.1 [192]) in three concentration levels. Due to the presence of endogenous amounts of a few calibrant lipids in platelets, RE, ME, and PE calculation was adjusted for endogenous concentrations of those lipids.

$$RE = \frac{C_{cal,pre}}{C_{cal,post}} * 100 \quad ME = \frac{C_{cal,post}}{C_{cal,neat}} * 100 \quad PE = \frac{C_{cal,pre}}{C_{cal,neat}} * 100 \quad (3.1)$$

A 9-point calibration curve plus unspiked blank matrix was used to evaluate the linear range for all 19 calibrant lipids corrected for endogenous concentrations if necessary. For the determination of accuracy and precision, three calibrants concentration levels ( $QC_{low}$ ,  $QC_{mid}$ ,  $QC_{high}$ ) were used with six injections per level on three individual days for intra- and inter-day calculations. LOD and LOQ were calculated as the concentrations with signal-to-noise (S/N) ratios of 3 and 10, respectively, based on the three lowest concentrations of the 9-point calibration curve. Due to the natural abundance of some calibrant lipids in platelets, LOD/LOQ determination was based on lipids peak height in neat solution calibration samples and multiplication of the obtained concentrations with the respective process efficiency.

### 3.2.1.8 Standard operating procedures (SOPs)

To ensure equal conditions in sampling, sample handling, sample preparation, and data acquisition, standard operating procedures (SOPs) were established. The documents contain

detailed information and descriptions of materials (instruments, lab ware, consumables, chemicals) and procedures. All SOPs are listed in Table 3.4. In addition to SOPs, Excel templates are also included which were used for calculations or label preparations.

Table 3.4: List of standard operating procedures established for the large-scale study

| <b>procedure</b>                  | <b>SOP</b>                                                                                                                              | <b>facility<sup>1</sup></b> |
|-----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|
| blood sampling                    | SOP blood sampling                                                                                                                      | clinic                      |
| platelet isolation                | SOP3 platelet isolation                                                                                                                 | clinic                      |
| lipid extraction                  | SOP_LipExtr_Lipidomics_4.docx<br><i>TR240-A08_Sample Randomization_1.xlsx</i><br><i>TR240-A08_lipid extract labels_2.xlsx</i>           | AK lab                      |
| internal standards<br>preparation | SOP_Prep IS solution_Lipidomics_1.docx<br><i>TR240-A08_Prep IS solution_Lipidomics_1.xlsx</i>                                           | AK lab                      |
| QC preparation                    | SOP_LipExtr_Lipidomics-QC_4.docx<br><i>TR240-A08_QC preparation_RP+HILIC_pos+neg.xlsx</i><br><i>TR240-A08_QC labels_RP+HILIC_2.xlsx</i> | AK lab                      |
| calibrants preparation            | SOP_Prep calibrant solution_Lipidomics_RP_1.docx<br><i>TR240-A08_Prep Calibrants_Lipidomics_3.xlsx</i>                                  | AK lab                      |
| system suitability test           | SOP_System suitability test_Lipidomics_RP_3.docx<br><i>TR240-A08_System suitability test_RP_evaluation.xlsx</i>                         | AK lab                      |
| LC-MS measurement                 | SOP_RP-LC-ESI-MS2_Lipidomics_4.docx<br><i>TR240-A08_Sequence preparation_RP_3.xlsx</i><br><i>TR240-A08_Mobile phase_RP_1.xlsx</i>       | AK lab                      |

<sup>1</sup> Facility that established the SOP and performed the corresponding procedures during the study.

### 3.2.2 Data processing and statistical analysis

The data processing procedure was developed and established during the entire project. The main focus in the prestudy data processing was the development of a suitable data extraction strategy for large-scale studies including data extraction, retention time adjustment, and inter-batch feature alignment. For the latter step, the development of an appropriate software tool was needed. Within the main study data processing, the processing procedure was refined and extended, especially regarding data curation of the reference feature list. The final data processing strategy for large-scale clinical lipidomics in platelets is described in the following.

#### 3.2.2.1 Quality control charts

To assess the proper quality of an acquired batch, quality control charts were generated. For each of the 19 internal standards, peak intensities of clinical and quality control samples were extracted from each batch using vendor software Sciex MultiQuant 3.0.3. (for parameters see Table A.6). The peak intensities were summed up, divided by the peak intensity of 18:1(d7) LPC, and plotted against the injection order. Based on the standard deviation (SD) from the mean sum peak intensity, respectively, warning ( $\pm 2\sigma$ ) and control levels ( $\pm 3\sigma$ ) were calculated and indicated in the QC chart. All batches from a cohort were combined in one

plot. The calculations for warning and control levels were performed batch-wise. QC charts were prepared separately for positive and negative ion mode.

### 3.2.2.2 Batch-wise untargeted data extraction

The initial step of the data processing strategy was the batch-wise untargeted data extraction of selected batches using open-source MS-DIAL software (RIKEN, version 2.94 (prestudy) and version 4.70 (main study)) [117]. MS-DIAL software was developed for processing DIA data like SWATH and offers automated peak finding, feature alignment, MS/MS spectra deconvolution, feature identification, and normalization methods.

The vendor data type (.wiff) was converted to *ABF*-format (Analysis Base File) and batches (three batches in prestudy and eight in main study) were processed using the parameters listed in Tables A.1 and A.2. Lipid identification was based on accurate mass and retention time similarity, isotope ratio, and MS/MS spectra matching of deconvoluted experimental with reference spectra from LipidBlast database [117]. Features were annotated as lipids with a total identification score  $>0.8$ . Data were extracted as peak heights, normalized with LOWESS normalization, and result tables were exported including PeakIDs, feature names,  $m/z$ -values, retention times, peak heights, and normalized peak heights of all detected features. Features with a coefficient of variation (CV) above 40% in QCs after LOWESS normalization were discarded from further processing.

### 3.2.2.3 Inter-batch feature alignment

Batch-wise data processing in MS-DIAL resulted in individual feature lists per batch and mode. To finally obtain a large-scale datamatrix including lipidomics data from multiple batches, inter-batch feature alignment had to be performed initially. Features detected in different batches had to be matched to obtain a representative reference feature list for the sample cohorts without redundancies. For this purpose, a suitable software tool was developed in Visual Basic for Applications (VBA) which was implemented in Excel. The tool design, alignment parameters, user instructions, and alignment performance are described in detail in section 4.4.

Prior feature alignment, retention time had to be adjusted for inter-batch retention time variations. Based on a set of internal standards and manually aligned features, covering the entire chromatographic range in each mode ( $\sim 2$  features/min), batch-specific retention times were linearly correlated with the mean retention time of the respective feature. The mean was calculated from three batches each in prestudy and in main study, respectively. Subsequently, retention times of all features of each batch and mode were adjusted to the mean retention time using the linear equation from correlation. For features eluting within a minute or less, the batch-specific retention times were kept.

Afterwards, the in-house developed tool was used for inter-batch feature alignment with feature lists from each acquisition mode and batch including PeakIDs,  $m/z$ -values, adjusted

retention times, feature names (identified lipids and unknown features), and peak heights in QC samples. Due to the tool design, the alignment was performed step-wise matching one batch against another batch (initially) or the actual reference feature list. Three parameters were required for the alignment and set to  $\Delta m/z = 0.01$ ,  $\Delta RT = 0.1$  min, and  $\Delta m/z_{closeness} = 0.016$  and  $\Delta m/z = 0.01$ ,  $\Delta RT = 0.15$  min, and  $\Delta m/z_{closeness} = 0.016$  for prestudy and main study data processing, respectively. For prestudy data, all three batches were used to prepare a reference feature list while eight batches were used in the main study (representing  $\sim 35\%$  of the sample cohort). In addition, the main study reference list was aligned with the identified lipids from the prestudy feature list. The general inter-batch alignment procedure is shown in Figure 3.2. Finally, reference feature lists were obtained for positive and negative ion mode in prestudy and main study data sets each.

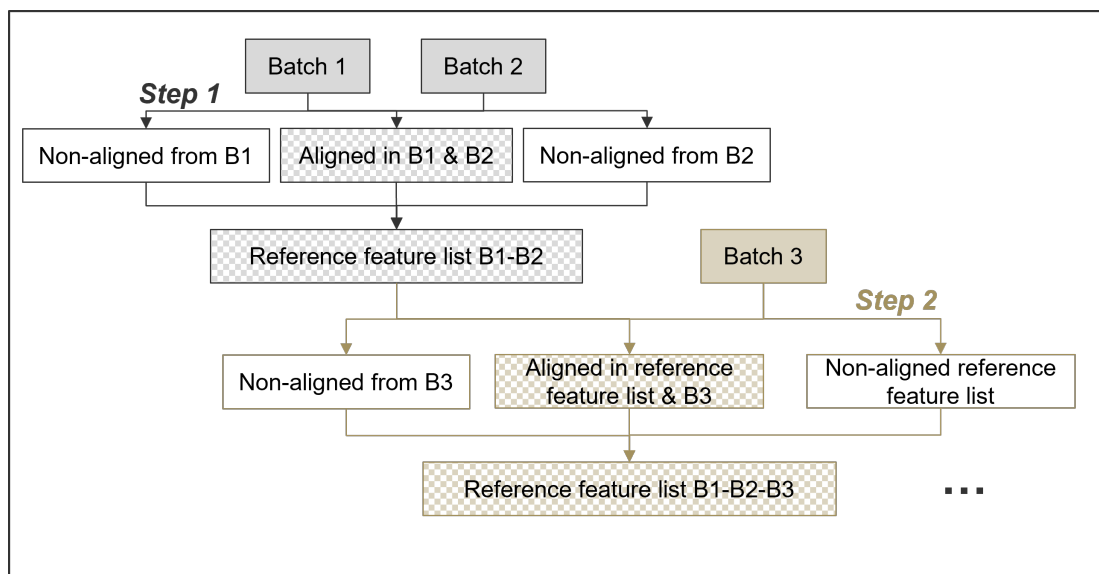


Figure 3.2: Scheme of inter-batch feature alignment procedure of multiple batches. In the first step, one batch was aligned against a second batch, while all following batches were aligned against the reference feature list from the former step.

#### 3.2.2.4 Data curation of reference feature list

Reference feature lists were further subjected to several data curation steps to confirm lipid identifications, detect and correct misidentifications, extension by lipids of interest, removal of redundancies (i.e. features detected in both positive and negative mode), and removal of false-positive features.

Firstly, automated lipid identification, performed in MS-DIAL data processing, was manually inspected for all features. To verify or reject lipid annotations, several parameters were taken into consideration: MS-DIAL identification score (score > 0.8), retention time information, i.e. matching the homologous series of the respective lipid class and adduct ion type, information from both ionization modes if existing. Lipids were annotated as recommended

by Liebisch et al. [46, 47]. For identified lipids being detected in both acquisition modes, merely the feature from the preferable mode was kept in the reference feature list in the main study.

The reference feature list from the main study was further extended by lipid species of special interest (species with medium chain fatty acyl(MCFA) side chains), lipid species to complete lipid class coverage (e.g. for FA, ACar lipid classes), or to include lipid species from a so far unconsidered lipid class (e.g. 3-hydroxy fatty acyls). For this purpose, targeted peak detection and identification based on accurate mass, retention time, and spectral information was performed, and detected features were added to the reference feature list.

Besides the identified features, data curation was also focused on unknown features. Unknown features might be lipids of unknown identity, e.g. unidentified due to missing reference spectra in database, due to noisy spectra, or due to too low abundance of the respective lipid for annotation. Several unknown features might represent the same lipid species because of lipids ionizing as multiple adduct types or in both positive and negative mode, or due to improper alignments. While redundant features from the two former cases were kept in the reference feature list, features from improper alignments were removed. For this purpose, unknown features of the reference list were investigated for accurate mass and retention time similarity within defined limits ( $\Delta m/z < 0.015$ ,  $\Delta RT < 0.15$  min) and removed after manual inspection if features represent the same chromatographic signal.

However, unknown features might also originate from noise or artefacts, with compromising effects on the statistical study outcome. Therefore, those false-positive features were removed using the open-source software EVA [193]. The tool was applied to all unknown features in the reference feature list for both acquisition modes separately. From each batch contributing to the reference feature list, a representative sample was converted to mzXML format in positive and negative mode and targeted data extraction was performed in MultiQuant to determine peaks start, end, and maximum retention time. After batch-wise application of EVA, features identified as false-positives in more than 50% of the samples, were discarded from the reference feature list.

After data curation, the final reference feature list was obtained. For each feature, the following information were summarized: 1. acquisition mode, 2. lipid species (sum), 3. lipid species with side chain annotation, 4. formula, 5. neutral exact mass, 6. adduct ion type, 7. precursor exact mass, 8. retention time, 9. lipid subclass, 10. lipid main class, 11. lipid category, 12. precursor accurate mass, 13. mass accuracy. Lipid classification was based on LIPID MAPS [46, 47].

### 3.2.2.5 Targeted data extraction and datamatrix preparation

Based on the finalized reference feature list, the target lists in positive and negative acquisition mode for batch-wise targeted data extraction were prepared. For targeted data extraction, a table containing the metabolite name, the precursor exact mass range, and the expected retention time were needed. The metabolite name for data extraction was modified by con-

catenation of information of lipid species (sum), lipid species with side chains, precursor ion type, and the entry number in the reference list. In addition to the precursor exact mass of a feature, additional precursor ions were included for sum extraction for some lipids. For internal standards, precursor exact masses of all detected adduct types as well as their isotopologues ([M-2], [M-1], [M+1], [M+2]) were considered. For Coenzyme Q10, a second precursor ion was included due to in-source reduction. The precursor ion mass extraction range was  $\pm 0.01$  Da of the exact mass.

Since the retention time in the reference feature list was the mean retention time, the batch-specific retention times were obtained by back calculation. An extended set of reference metabolites in contrast to the retention time adjustment prior inter-batch alignment (41 and 31 metabolites in positive and negative mode, respectively; max.  $\Delta RT_{\text{metabolites}} < 0.5$  min) was used for retention time adjustment based on linear correlation.

The final target lists were subjected to targeted data extraction using MultiQuant (parameters are listed in Table A.6). Data from each batch and acquisition mode were extracted separately as peak heights and the lipidomics data of all batches were then combined in one datamatrix per acquisition mode.

The datamatrix was further evaluated in terms of feature detection rates and proportion of missing values. Based on the detection rates, features not being detected in all batches were discarded. By determination of missing values proportion, samples with  $>80\%$  missing values were removed. Afterwards, missing values were imputed by the half-minimum method (prestudy) or random forest (RF) algorithm (main study) using the package *missForest* [194] in RStudio (R version 4.2.0.) [195]. Random forest is a non-parametric approach well suitable for untargeted MS metabolomics data [194, 196]. Moreover, for all identified lipids being present as multiple precursor ions, peak heights of all detected adducts were summed up to a single datamatrix entry.

Finally, type I isotopic correction for carbon isotopes was performed using the Equation 3.2 based on Wang et al. [86] and Lange et al. [156].

$$intensity_{n(k)total} = intensity_{n(k)} \left( 1 + 0.0109n + \frac{0.0109^2 n(n-1)}{2} \right) \quad (3.2)$$

Thereby,  $n$  is the number of carbon atoms and  $k$  represents the number of double bonds. Further, the peak intensity is used instead of peak area.

### 3.2.2.6 Optimization and application of data normalization

Once the raw lipidomics data were consolidated and curated in acquisition-mode-wise datamatrices, data normalization was performed. Normalization was essential due to the large-scale study design with several acquisition batches acquired over a long time period. A large diversity of normalization approaches and respective bioinformatic solutions were available either

based on sample-specific scaling factor, internal standards, or quality control samples. To select the optimum normalization strategy for the pre- and main study data sets, multiple approaches were applied and their performance was evaluated.

Data normalization and performance evaluation was performed in RStudio (R version 3.5.2, 3.6.0, 4.2.0 depending on the required packages) using *metabolomics* [197], *MetNorm* [198], and *NormalizeMets* [199] packages based on published R scripts from Drotleff et al. [180] and Fan et al. [159].

In the prestudy data set, all three batches were used to test for the optimum normalization method comparing the performance of IS-based CCMN, NOMIS, B-MIS, an in-house developed IS-based normalization, and RUVrandom, QC-based methods LOWESS, and SERRF, as well as a combination of SERRF and RUVrandom. The optimum normalization method for the main study data set was determined based on eight batches (same as used for reference feature list preparation). Twelve approaches were compared: IS-based methods CCMN, NOMIS, B-MIS, and RUVrandom, QC-based methods LOESS, SERRF, and SVM, sample-based methods PCnorm, PQN, mTIC, as well as combinations of SERRF with CCMN and mTIC with CCMN.

The normalization performance was evaluated by uni- and multivariate data analysis (UVDA and MVDA) and variation considerations following the recommendations from Drotleff et al. [180]. Based on partial least squares discriminant analysis (PLS-DA) and assessment of variation by group-dependent coefficient of variation (CV), median absolute deviation (MAD), and variance, the optimum normalization method was chosen and applied to the entire data set.

### 3.2.2.7 Final datamatrix preparation

As the last step, the final datamatrix was prepared. Firstly, features with a high variation were removed based on the FDA's precision criteria in biomarker studies. Thus, the inter-batch CV for normalized peak height in QCs was calculated and features with a CV > 30% were discarded from raw and normalized datamatrices.

Furthermore, lipid concentrations of identified lipids were calculated in pmol/10<sup>6</sup> cells by one-point calibration with the lipid class specific isotopically labeled internal standard.

Finally, lipidomics data from positive and negative acquisition mode of raw and normalized data as well as lipid concentration data of identified lipids were combined in three respective matrices and linked with clinical sample information. As a summary, all processing steps performed in prestudy and main study data sets are listed in Table 3.5.

Table 3.5: Steps of lipidomics data processing performed in pre- and main study data sets with three and 22 acquisition batches, respectively

| processing step                                             | prestudy                | main study                                                     |
|-------------------------------------------------------------|-------------------------|----------------------------------------------------------------|
| Quality control charts                                      | 3 batches               | 22 batches                                                     |
| MS-DIAL data processing                                     | 3 batches               | 8 batches                                                      |
| Inter-batch feature alignment                               | 3 batches               | 8 batches                                                      |
| Curation of identified lipids                               | yes                     | yes                                                            |
| Removal of redundant lipids<br>(acquisition mode selection) | no                      | yes                                                            |
| Manual addition of lipid targets                            | no                      | yes                                                            |
| Curation of unknown features                                | no                      | yes                                                            |
| Targeted data extraction                                    | 3 batches               | 22 batches                                                     |
| Missing values imputation                                   | RF                      | RF                                                             |
| Totalization of multiple adducts                            | no                      | yes                                                            |
| Type I isotopic correction                                  | no                      | per calculation, exception:<br>isotopologues extraction for IS |
| Data normalization                                          | SERRF-RUVrandom         | SERRF                                                          |
| Removal of features                                         | $CV_{QC, norm} > 30\%$  | $CV_{QC, norm} > 30\%$                                         |
| Lipid concentrations                                        | lipid class specific IS | lipid class specific IS                                        |

### 3.2.2.8 Statistical analysis

The final lipidomics datamatrix was subjected to statistical analysis. Univariate data analysis was performed in RStudio (R version 3.6.0) with *metabolomics* [197] and *MetNorm* [198] packages and JMP<sup>®</sup> (version 14.2.0, SAS Institute, Cary, NC, USA). In detail, normality and variance homogeneity testing with Shapiro-Wilk test and Bartlett test, respectively, as well as non-parametric Mann-Whitney-U-test and parametric Student's t-test and one-way ANOVA were performed. For multivariate analysis, i.e. PCA, PLS-DA, OPLS-DA, SIMCA<sup>®</sup> from Umetrics (version 16.0, Sartorius AG, Göttingen, Germany) and again RStudio were used.



## 4 Results and discussion

### 4.1 Method validation

The LC-MS/MS method applied in this large-scale clinical lipidomics study of platelets was initially developed and published by Tsugawa et al. [117]. In our laboratory, the method has been used frequently in lipidomics investigations on different cells (keratinocytes [66], HeLa [191], platelets [200]) and plasma [85, 180]. However, the method was validated for the current study setting to assess proper analytical performance. For this purpose, authentic standards and their respective stable labeled internal standards were spiked to an identical pool of platelets. Adapted from FDA validation guidelines for targeted assays [146], accuracy, precision, linear ranges, limits of detection (LOD) and quantification (LOQ) were determined. In addition, recovery (RE), matrix effect (ME), and process efficiency (PE) were evaluated following the recommendations from Matuszewski et al. [192].

For validation, a set of standards was chosen which matched the spiked internal standards for the clinical study samples. Since the commercial internal standard mixture EquiSPLASH® LIPIDOMIX®, added by a few more IS to cover all investigated lipid classes (see section 3.2.1.3), was used, the corresponding primary standards mixture LightSPLASH® LIPIDOMIX®, similarly complemented with a few more standards like the IS mixture, was selected for validation. To obtain an identical sample matrix for all validation samples, a platelet pool was prepared by combining 16 platelet pellets with a total count of  $117.65 \times 10^8$ . Pellets were re-suspended in PBS buffer, pooled, and aliquoted again to  $1.5 \times 10^8$  (QC samples) and  $2.0 \times 10^8$  cells per aliquot (calibration curve preparation). Due to the limited sample availability, a lower cell count had to be used in QC samples to be able to prepare all required validation samples. While QC samples were used for intra- and inter-batch accuracy and precision calculations, calibration samples were used for linear range, calibration, LOD, and LOQ determination.

The validation results regarding linear ranges and calibration, LOD, LOQ as well as accuracy and precision are summarized in Tables 4.1 and 4.2. For some calibrants, significant endogenous lipid amounts were observed, e.g. 15:0-18:1 PC (219 ng/mL), d18:1-18:1 SM (276 ng/mL), Cholesterol (38267 ng/mL), and 18:1 Cholesteryl Ester (244 ng/mL). The endogenous concentration levels were determined by extrapolating the calibration curve to the abscissa. To account for the matrix lipid concentrations, spiked concentration levels were added by endogenous lipid amounts for the calibration results as well as precision and accuracy calculations (standard addition approach).

Table 4.1: Summary of validation results showing the linear range, calibration parameters, as well as LOD and LOQ (quantified with respective IS). LOD and LOQ were determined in neat solution as  $S/N \approx 3$  and  $S/N \approx 10$ , respectively, multiplied by the calibrant specific process efficiency. Further results were obtained under consideration of endogenous amounts of the calibrant lipids

| mode | calibrant         | RT<br>[min] | linear range<br>[ng/mL] | slope  | calibration<br>y-intercept | $R^2$  | weighting | LOD [ng/mL] | LOQ [ng/mL] |
|------|-------------------|-------------|-------------------------|--------|----------------------------|--------|-----------|-------------|-------------|
| pos  | 15:0-18:1 PC      | 5.15        | 0.5-30000               | 1.0990 | -7.9E-02                   | 0.9989 | 1/x       | 0.5         | 1.6         |
|      | 18:1 LPC          | 1.57        | 0.4-30000               | 1.1984 | -1.5E-02                   | 0.9987 | 1/x       | 0.4         | 1.4         |
|      | 15:0-18:1 PE      | 5.32        | 3.0-30000               | 1.1222 | -3.3E-02                   | 0.9951 | 1/x       | 3.0         | 10.0        |
|      | 18:1 LPE          | 1.61        | 3.6-30000               | 1.1485 | -9.5E-03                   | 0.9932 | 1/x       | 3.6         | 12.0        |
|      | 15:0-18:1 PG      | 4.67        | 5.2-30000               | 0.9892 | -4.0E-04                   | 0.9986 | 1/x       | 5.2         | 17.5        |
|      | 15:0-18:1 PI      | 4.48        | 17.4-30000              | 1.0941 | 8.1E-05                    | 0.9971 | 1/x       | 17.4        | 58.1        |
|      | d18:1-18:1 SM     | 4.77        | 0.9-30000               | 0.9685 | -6.7E-02                   | 0.9972 | 1/x       | 0.9         | 2.9         |
|      | LysoSM d18:1      | 0.91        | 1.5-30000               | 0.4100 | -1.2E-02                   | 0.9964 | 1/x       | 1.5         | 5.1         |
|      | C15 Ceramide      | 5.21        | 3.8-30000               | 0.6252 | 1.9E-03                    | 0.9953 | 1/x       | 3.8         | 12.6        |
|      | L-CAR 16:0        | 1.56        | 0.1-30000               | 0.8451 | -1.5E-02                   | 0.9965 | 1/x       | 0.1         | 0.4         |
|      | Cholesterol       | 4.75        | 6.8-46600               | 1.0314 | 4.1E-01                    | 0.8297 | 1/x       | 6.8         | 22.7        |
|      | 18:1 Chol Ester   | 10.89       | 28.1-500000             | 1.3507 | -1.8E-02                   | 0.9989 | 1/x       | 28.1        | 93.6        |
|      | 18:1 MG           | 2.73        | 6.1-30000               | 0.9587 | -1.4E-03                   | 0.9979 | 1/x       | 6.1         | 20.2        |
|      | 15:0-18:1 DG      | 6.53        | 2.5-30000               | 1.1360 | -7.4E-03                   | 0.9976 | 1/x       | 2.5         | 8.4         |
|      | 15:0-18:1-15:0 TG | 10.42       | 0.7-20000               | 1.2498 | -8.1E-02                   | 0.9986 | 1/x       | 0.7         | 2.2         |
| neg  | 15:0-18:1 PC      | 5.15        | 0.8-30000               | 1.1422 | -8.2E-02                   | 0.9993 | 1/x       | 0.8         | 2.7         |
|      | 18:1 LPC          | 1.57        | 1.1-30000               | 1.2508 | 3.9E-03                    | 0.9953 | 1/x       | 1.1         | 3.7         |
|      | 15:0-18:1 PE      | 5.32        | 0.5-30000               | 0.9973 | -5.5E-03                   | 0.9971 | 1/x       | 0.5         | 1.8         |
|      | 18:1 LPE          | 1.61        | 1.0-30000               | 1.1559 | 1.3E-02                    | 0.9945 | 1/x       | 1.0         | 3.2         |
|      | 15:0-18:1 PG      | 4.67        | 0.7-30000               | 1.0710 | -1.9E-02                   | 0.9972 | 1/x       | 0.7         | 2.5         |
|      | 15:0-18:1 PI      | 4.48        | 3.2-30000               | 1.0199 | -1.3E-02                   | 0.9978 | 1/x       | 3.2         | 10.5        |
|      | 15:0-18:1 PS      | 4.58        | 4.9-30000               | 1.0635 | -2.5E-02                   | 0.9963 | 1/x       | 4.9         | 16.3        |
|      | d18:1-18:1 SM     | 4.77        | 3.0-30000               | 1.0176 | -8.2E-02                   | 0.9956 | 1/x       | 3.0         | 10.0        |
|      | LysoSM d18:1      | 0.91        | 2.0-30000               | 0.6139 | -4.6E-03                   | 0.9906 | 1/x       | 2.0         | 6.5         |
|      | C15 Ceramide      | 5.21        | 0.4-30000               | 0.6894 | -4.6E-03                   | 0.9937 | 1/x       | 0.4         | 1.3         |
|      | L-CAR 16:0        | 1.56        | 1.4-30000               | 0.9985 | 4.3E-03                    | 0.9967 | 1/x       | 1.4         | 4.6         |
|      | 18:1 MG           | 2.73        | 3.5-30000               | 1.1536 | -2.1E-02                   | 0.9954 | 1/x       | 3.5         | 11.7        |
|      | 20:4 AA           | 2.65        | 2.5-30000               | 0.6198 | 6.4E-02                    | 0.9918 | 1/x       | 2.5         | 8.3         |

Table 4.2: Summary of validation results for within-batch and between-batch (i.e. three different days) precision and accuracy in QC samples spiked with three concentration levels of calibrants. Results were obtained under consideration of endogenous amounts of the calibrant lipids

| mode | calibrant         | precision [%]     |                   |                    |                   |                   |                    | accuracy [%]      |                   |                    |                   |                   |                    |
|------|-------------------|-------------------|-------------------|--------------------|-------------------|-------------------|--------------------|-------------------|-------------------|--------------------|-------------------|-------------------|--------------------|
|      |                   | within-batch      |                   |                    | between-batch     |                   |                    | within-batch      |                   |                    | between-batch     |                   |                    |
|      |                   | QC <sub>low</sub> | QC <sub>mid</sub> | QC <sub>high</sub> | QC <sub>low</sub> | QC <sub>mid</sub> | QC <sub>high</sub> | QC <sub>low</sub> | QC <sub>mid</sub> | QC <sub>high</sub> | QC <sub>low</sub> | QC <sub>mid</sub> | QC <sub>high</sub> |
| pos  | 15:0-18:1 PC      | 5.6               | 3.6               | 2.3                | 3.9               | 3.0               | 1.8                | 100.5             | 96.4              | 100.3              | 100.2             | 97.8              | 100.2              |
|      | 18:1 LPC          | 5.7               | 3.3               | 4.8                | 6.0               | 3.6               | 5.4                | 100.3             | 97.6              | 100.2              | 99.1              | 94.9              | 100.5              |
|      | 15:0-18:1 PE      | 11.7              | 7.1               | 5.9                | 11.3              | 5.7               | 5.2                | 94.5              | 97.0              | 100.4              | 95.6              | 94.4              | 100.6              |
|      | 18:1 LPE          | 4.3               | 7.1               | 9.1                | 4.9               | 4.4               | 7.6                | 96.5              | 97.4              | 100.3              | 97.2              | 97.2              | 100.3              |
|      | 15:0-18:1 PG      | 12.7              | 6.7               | 7.3                | 13.5              | 7.8               | 7.9                | 100.4             | 106.8             | 99.3               | 98.3              | 103.5             | 99.7               |
|      | 15:0-18:1 PI      | 7.5               | 9.3               | 3.5                | 5.8               | 6.9               | 6.7                | 92.5              | 91.5              | 114.3              | 90.6              | 88.6              | 106.2              |
|      | d18:1-18:1 SM     | 7.4               | 4.3               | 2.7                | 5.7               | 5.6               | 4.8                | 95.0              | 103.7             | 100.0              | 97.0              | 103.6             | 99.8               |
|      | LysoSM d18:1      | 14.7              | 13.3              | 20.7               | 10.8              | 11.5              | 16.1               | 98.8              | 110.2             | 99.0               | 96.3              | 106.0             | 99.4               |
|      | C15 Ceramide      | 9.1               | 9.5               | 8.2                | 7.7               | 7.6               | 9.4                | 97.4              | 99.0              | 100.1              | 97.9              | 103.1             | 99.7               |
|      | L-CAR 16:0        | 3.0               | 2.6               | 9.8                | 4.1               | 3.4               | 8.6                | 100.6             | 108.5             | 91.4               | 101.6             | 110.7             | 96.0               |
|      | Cholesterol       | 6.3               | 18.1              | 29.7               | 7.8               | 17.0              | 19.9               | 97.2              | 106.4             | 99.4               | 100.6             | 98.4              | 100.1              |
|      | 18:1 Chol Ester   | 3.6               | 4.3               | 6.1                | 6.1               | 6.9               | 5.0                | 104.6             | 106.0             | 99.4               | 102.6             | 106.7             | 99.3               |
|      | 18:1 MG           | 11.2              | 4.0               | 4.0                | 9.2               | 6.2               | 6.7                | 102.0             | 101.6             | 99.8               | 99.6              | 97.7              | 100.2              |
|      | 15:0-18:1 DG      | 3.1               | 6.0               | 5.2                | 5.3               | 4.3               | 4.9                | 94.7              | 94.5              | 100.6              | 93.3              | 93.8              | 100.7              |
|      | 15:0-18:1-15:0 TG | 5.1               | 4.5               | 2.8                | 5.7               | 6.4               | 3.9                | 95.7              | 85.2              | 101.5              | 92.6              | 90.2              | 101.0              |
| neg  | 15:0-18:1 PC      | 10.8              | 2.3               | 3.3                | 8.9               | 2.8               | 3.6                | 97.9              | 97.3              | 100.3              | 99.5              | 95.9              | 100.4              |
|      | 18:1 LPC          | 3.5               | 3.4               | 2.4                | 2.8               | 2.9               | 3.3                | 93.8              | 92.2              | 100.8              | 93.4              | 92.4              | 100.8              |
|      | 15:0-18:1 PE      | 4.0               | 2.5               | 4.1                | 6.3               | 3.4               | 4.0                | 93.8              | 90.7              | 101.0              | 91.3              | 90.9              | 101.0              |
|      | 18:1 LPE          | 4.9               | 2.2               | 4.5                | 5.2               | 2.0               | 3.5                | 96.6              | 97.0              | 100.3              | 98.4              | 97.1              | 100.3              |
|      | 15:0-18:1 PG      | 7.4               | 7.3               | 3.7                | 6.9               | 5.4               | 3.5                | 95.4              | 89.7              | 101.1              | 92.8              | 88.6              | 101.2              |
|      | 15:0-18:1 PI      | 7.6               | 2.1               | 4.5                | 7.0               | 3.0               | 4.2                | 92.7              | 91.8              | 100.9              | 92.0              | 93.3              | 100.7              |
|      | 15:0-18:1 PS      | 6.4               | 3.3               | 4.7                | 9.2               | 6.2               | 4.1                | 89.9              | 92.7              | 100.8              | 87.2              | 91.6              | 100.9              |
|      | d18:1-18:1 SM     | 8.0               | 6.0               | 5.9                | 6.3               | 4.8               | 8.1                | 99.3              | 94.1              | 100.6              | 100.5             | 96.0              | 100.4              |
|      | LysoSM d18:1      | 17.4              | 14.6              | 12.8               | 21.5              | 16.0              | 18.9               | 98.8              | 93.8              | 100.7              | 92.1              | 92.6              | 105.5              |
|      | C15 Ceramide      | 16.6              | 8.3               | 14.6               | 16.2              | 10.8              | 11.3               | 91.1              | 84.0              | 101.7              | 90.2              | 90.0              | 101.1              |
|      | L-CAR 16:0        | 6.5               | 2.7               | 7.8                | 5.7               | 2.8               | 6.0                | 94.5              | 96.4              | 100.4              | 94.0              | 97.7              | 100.3              |
|      | 18:1 MG           | 9.1               | 9.0               | 5.3                | 7.2               | 6.3               | 4.5                | 95.8              | 94.5              | 100.6              | 95.8              | 96.3              | 100.4              |
|      | 20:4 AA           | 16.7              | 6.9               | 11.7               | 13.4              | 9.9               | 9.7                | 91.1              | 95.8              | 100.5              | 95.7              | 105.3             | 99.5               |

The calibration samples were prepared by spiking primary standards (=calibrants) at nine different concentration levels and constant concentrations of internal standard mixture to the platelet matrix prior extraction. The calibration series was prepared in replicates for inter-day measurements. After adjusting the calibration results for endogenous lipid concentrations (Table 4.1), a good linearity over the entire tested concentration range was obtained for all lipid standards except of Cholesterol ( $R^2=0.8297$ ) due to its high levels in platelets. In future, it might be reasonable to replace it by a surrogate like a  $^{13}\text{C}$ -labeled standard. Due to endogenous lipid levels of some calibrants, LOD and LOQ were determined as  $S/N\approx 3$  and  $S/N\approx 10$ , respectively, in a calibration series prepared in neat solution (=methanol). To account for matrix effects and extraction recoveries, the obtained concentrations were multiplied by the calibrant specific process efficiency determined in  $\text{QC}_{\text{low}}$  samples (Table 4.3).

Accuracy and precision were determined in QC samples with low, mid, and high calibrants concentrations prepared in replicates for intra-day as well as inter-day measurements (Table 4.2). The results were also adjusted for endogenous concentrations. The validation results indicate good performance in terms of precision and accuracy within a batch as well as between batches acquired on different days. The majority of calibrants in all three concentration levels passes the acceptance criteria of  $\pm 15\%$  usually applied for targeted assays. For Cholesterol, the precision is declining with increasing spiked concentration, caused by already high lipid levels in matrix. However, in accordance with former studies [66, 85, 180, 191, 200], the validation results reveal a good and robust performance of the applied LC-MS/MS method.

For the determination of recovery, matrix effect, and process efficiency, QC samples were pre-extraction and post-extraction spiked with calibrants at three different concentration levels while concentrations of the respective internal standards were kept constantly. Similarly, calibrants and internal standards were prepared in methanol as neat solution. The results are summarized in Table 4.3. After adjusting for endogenous amounts of some calibrant lipids as mentioned above, extraction recoveries were mainly closely around 100% with exceptions of Cholesterol (135.2%) and TG (111.0%). Only a limited extent of matrix effects are observed for the calibrants, usually due to ion suppression (i.e.  $\text{ME} < 100\%$ ). Consequently, process efficiencies for the investigated lipids is generally optimal in the range of  $100\% \pm 15\%$  (except of PS with 84%) indicating a sufficient method performance for the relevant lipid classes of interest.

Table 4.3: Summary of validation results for recovery (RE), matrix effect (ME), and process efficiency (PE) for calibrants calculated as the mean $\pm$ standard deviation for three concentration levels

| mode | calibrant                | RE [%]           | ME [%]          | PE [%]           |
|------|--------------------------|------------------|-----------------|------------------|
| pos  | <b>15:0-18:1 PC</b>      | 100.4 $\pm$ 1.3  | 104.5 $\pm$ 2.1 | 104.9 $\pm$ 2.0  |
|      | <b>18:1 LPC</b>          | 98.9 $\pm$ 5.1   | 96.1 $\pm$ 2.6  | 94.9 $\pm$ 2.3   |
|      | <b>15:0-18:1 PE</b>      | 99.5 $\pm$ 3.7   | 97.4 $\pm$ 3.6  | 96.9 $\pm$ 4.8   |
|      | <b>18:1 LPE</b>          | 100.0 $\pm$ 3.4  | 104.2 $\pm$ 5.3 | 104.0 $\pm$ 1.7  |
|      | <b>15:0-18:1 PG</b>      | 100.1 $\pm$ 1.5  | 88.7 $\pm$ 4.2  | 88.8 $\pm$ 3.0   |
|      | <b>15:0-18:1 PI</b>      | 99.6 $\pm$ 2.0   | 93.1 $\pm$ 7.5  | 92.8 $\pm$ 8.3   |
|      | <b>d18:1-18:1 SM</b>     | 99.9 $\pm$ 0.9   | 109.3 $\pm$ 2.7 | 109.1 $\pm$ 2.7  |
|      | <b>LysoSM d18:1</b>      | 99.2 $\pm$ 6.9   | 106.6 $\pm$ 8.4 | 105.8 $\pm$ 11.3 |
|      | <b>C15 Ceramide</b>      | 99.5 $\pm$ 3.9   | 101.4 $\pm$ 2.2 | 100.8 $\pm$ 1.9  |
|      | <b>L-CAR 16:0</b>        | 102.3 $\pm$ 4.1  | 98.8 $\pm$ 6.8  | 101.0 $\pm$ 7.5  |
|      | <b>Cholesterol</b>       | 135.2 $\pm$ 9.7  | 73.5 $\pm$ 3.1  | 99.1 $\pm$ 4.1   |
|      | <b>18:1 Chol Ester</b>   | 106.9 $\pm$ 6.3  | 96.6 $\pm$ 5.1  | 103.0 $\pm$ 2.7  |
|      | <b>18:1 MG</b>           | 102.9 $\pm$ 5.4  | 89.1 $\pm$ 2.6  | 91.5 $\pm$ 2.1   |
|      | <b>15:0-18:1 DG</b>      | 101.6 $\pm$ 3.6  | 93.4 $\pm$ 5.4  | 94.7 $\pm$ 3.6   |
|      | <b>15:0-18:1-15:0 TG</b> | 111.0 $\pm$ 17.9 | 84.7 $\pm$ 16.4 | 91.5 $\pm$ 9.7   |
| neg  | <b>15:0-18:1 PC</b>      | 100.2 $\pm$ 3.3  | 105.9 $\pm$ 3.1 | 106.0 $\pm$ 1.8  |
|      | <b>18:1 LPC</b>          | 98.6 $\pm$ 0.8   | 97.3 $\pm$ 3.6  | 95.9 $\pm$ 3.3   |
|      | <b>15:0-18:1 PE</b>      | 101.8 $\pm$ 1.6  | 86.1 $\pm$ 2.9  | 87.7 $\pm$ 4.4   |
|      | <b>18:1 LPE</b>          | 100.3 $\pm$ 1.6  | 101.6 $\pm$ 0.3 | 102.0 $\pm$ 1.3  |
|      | <b>15:0-18:1 PG</b>      | 100.4 $\pm$ 2.5  | 85.5 $\pm$ 4.5  | 85.8 $\pm$ 4.8   |
|      | <b>15:0-18:1 PI</b>      | 102.3 $\pm$ 1.1  | 89.8 $\pm$ 2.5  | 91.9 $\pm$ 3.3   |
|      | <b>15:0-18:1 PS</b>      | 101.4 $\pm$ 1.0  | 82.9 $\pm$ 5.1  | 84.0 $\pm$ 4.7   |
|      | <b>d18:1-18:1 SM</b>     | 98.2 $\pm$ 3.9   | 104.4 $\pm$ 4.7 | 102.4 $\pm$ 3.2  |
|      | <b>LysoSM d18:1</b>      | 107.1 $\pm$ 7.6  | 92.2 $\pm$ 4.4  | 99.0 $\pm$ 11.7  |
|      | <b>C15 Ceramide</b>      | 101.6 $\pm$ 3.9  | 90.3 $\pm$ 5.3  | 91.9 $\pm$ 9.0   |
|      | <b>L-CAR 16:0</b>        | 100.3 $\pm$ 0.7  | 98.2 $\pm$ 2.0  | 98.5 $\pm$ 2.6   |
|      | <b>18:1 MG</b>           | 99.5 $\pm$ 2.7   | 100.8 $\pm$ 0.1 | 100.3 $\pm$ 2.8  |
|      | <b>20:4 AA</b>           | 102.4 $\pm$ 4.9  | 97.7 $\pm$ 1.6  | 100.0 $\pm$ 4.1  |

## 4.2 Quality control

Prior data processing, batch performance, i.e. quality control, was assessed based on spiked internal standards. For this purpose, the coefficient of variation (CV) was determined in QC samples which were injected regularly across each batch sequence. In addition, QC charts as a visual representation were prepared.

The coefficient of variation for internal standards in QC samples was calculated from non-normalized, raw peak intensities and the results are summarized in Table 4.4 for the prestudy measurements and in Figure 4.2 for the main study cohort. For the prestudy dataset (Table 4.4), the CVs indicate stable, reproducible batch acquisition in both positive and negative mode based on CVs below 15% for the majority of IS. Only the IS of PE and Chol Ester have

a slightly higher variation with mean CV values across the three batches of 18.6% and 17.6%, respectively.

Table 4.4: Coefficient of variation of internal standards determined from raw peak intensities in QC samples regularly injected over the sequences of prestudy batches; mean CV with standard deviation over all three batches is also given

| mode | internal standard     | coefficient of variation [%] |          |          |                |
|------|-----------------------|------------------------------|----------|----------|----------------|
|      |                       | Batch 01                     | Batch 02 | Batch 03 | mean $\pm$ SD  |
| pos  | 15:0-18:1(d7) PC      | 2.7                          | 6.6      | 4.0      | 4.4 $\pm$ 1.6  |
|      | 18:1(d7) LPC          | 4.2                          | 7.4      | 6.1      | 5.9 $\pm$ 1.3  |
|      | 15:0-18:1(d7) PE      | 19.3                         | 14.1     | 22.5     | 18.6 $\pm$ 3.5 |
|      | 18:1(d7) LPE          | 7.0                          | 3.8      | 11.1     | 7.3 $\pm$ 3.0  |
|      | 15:0-18:1(d7) PI      | 11.2                         | 13.1     | 13.0     | 12.4 $\pm$ 0.8 |
|      | d18:1-18:1(d9) SM     | 10.8                         | 12.9     | 9.8      | 11.1 $\pm$ 1.3 |
|      | d18:1-18:0(d7) Cer    | 11.0                         | 13.3     | 14.6     | 13.0 $\pm$ 1.5 |
|      | 18:1(d7) Chol Ester   | 19.0                         | 18.9     | 14.9     | 17.6 $\pm$ 1.9 |
|      | 15:0-18:1(d7) DG      | 6.5                          | 8.1      | 7.9      | 7.5 $\pm$ 0.7  |
|      | 15:0-18:1(d7)-15:0 TG | 6.5                          | 10.2     | 12.3     | 9.7 $\pm$ 2.4  |
| neg  | 15:0-18:1(d7) PC      | 5.8                          | 6.9      | 5.9      | 6.2 $\pm$ 0.5  |
|      | 18:1(d7) LPC          | 3.3                          | 4.3      | 12.3     | 6.6 $\pm$ 4.0  |
|      | 18:1(d7) LPE          | 6.1                          | 4.7      | 12.4     | 7.7 $\pm$ 3.3  |
|      | 15:0-18:1(d7) PG      | 10.8                         | 12.8     | 11.5     | 11.7 $\pm$ 0.9 |
|      | d18:1-18:1(d9) SM     | 11.3                         | 7.9      | 13.4     | 10.9 $\pm$ 2.2 |
|      | d18:1-18:0(d7) Cer    | 6.9                          | 7.5      | 10.9     | 8.4 $\pm$ 1.8  |
|      | 20:4(d8) AA           | 11.8                         | 11.4     | 9.5      | 10.9 $\pm$ 1.0 |

Further quality assessment was based on quality control charts, for which the raw peak intensities of all internal standards of the respective ionization mode were summed up, normalized by the peak intensity of the LPC-IS (18:1(d7) LPC), and plotted against the injection order over all three batches. In addition, warning ( $\pm 2\sigma$ ) and control ( $\pm 3\sigma$ ) levels were included to support the visual assessment of data quality. Usually, quality control charts are prepared for each target analyte individually. In this untargeted approach, QC charts are based on peak intensities of internal standards, and for a neat batch representation, all detected IS are totalized to be plotted in one figure instead of individual plots for each IS.

The prestudy QC charts shown in Figure 4.1 confirm the proper measurement performance in terms of intra- and inter-batch stability. Two samples in negative ion mode were outside the  $\pm 3\sigma$ -range. Since the positive mode of those samples performed well and caused by the limited sample matrix, no remeasurement was performed but the fluctuations were addressed by post-acquisitional normalization.

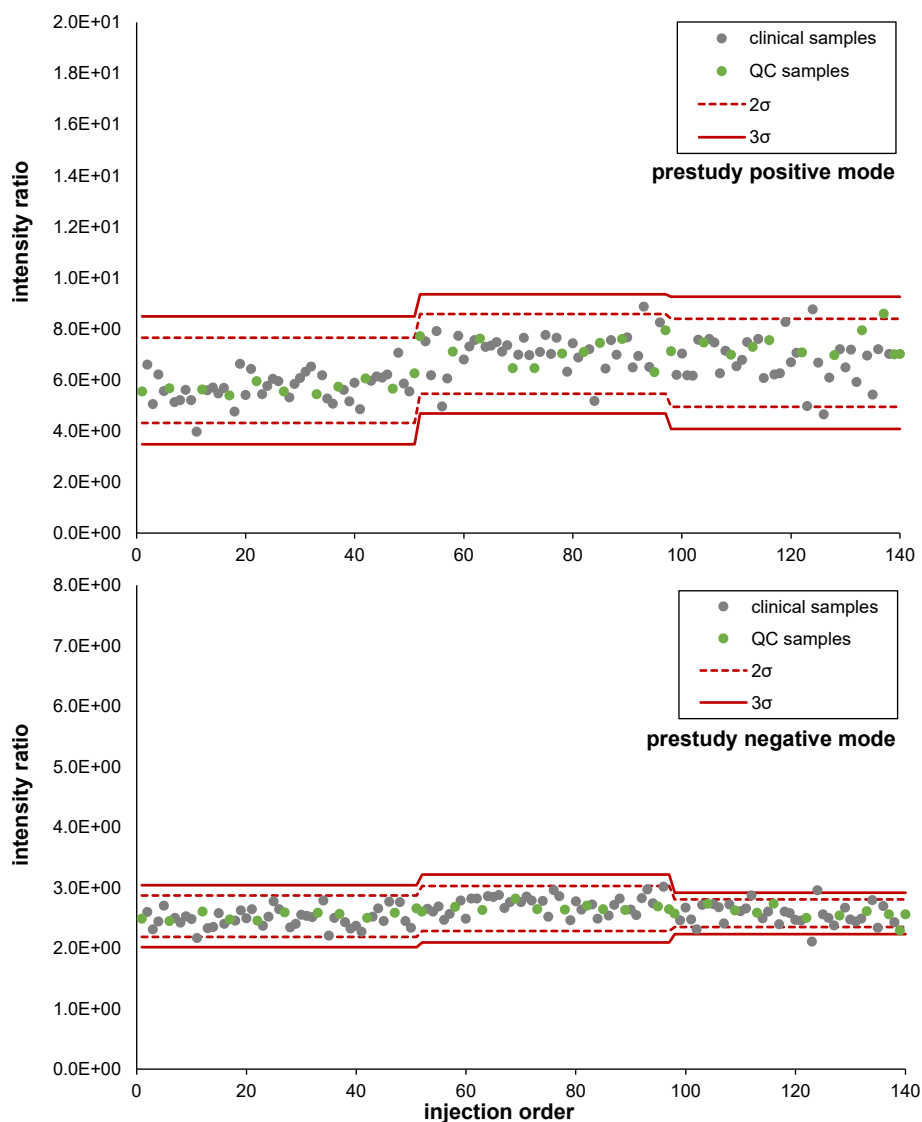


Figure 4.1: Prestudy quality control charts showing peak intensity ratios of internal standards in clinical and QC samples plotted against the injection order over all three acquisition batches in positive and negative ionization mode each. Depicted intensity ratios were obtained by referring the sum peak intensities of all internal standards in the respective ionization mode (pos: 10 IS; neg: 7 IS) to the peak intensity of 18:1(d7) LPC. Batch-wise calculated  $\pm 2\sigma$  and  $\pm 3\sigma$  from the mean indicate warning and control levels, respectively.

A different representation of the CV values was chosen for the main study cohort due to the larger number of IS and batches. Thus, CV values were calculated for each IS in QC samples and a median CV was then determined representing the variation within a batch and ionization mode (Figure 4.2; positive ion mode: 15 IS; negative ion mode: 13 IS)). The results in Figure 4.2 indicate low variation in the majority of the acquired batches. However, a few batches (B06, B08, B10, B12, B13) show higher variation in positive mode which is caused by intensity shifts during the sequence. Since the shifts were constantly observed for all IS in

the QC samples, but was not a sample inherent variation (the negative mode variation for the respective batches was much lower), the measurement fluctuations can be addressed properly by suitable normalization strategies.

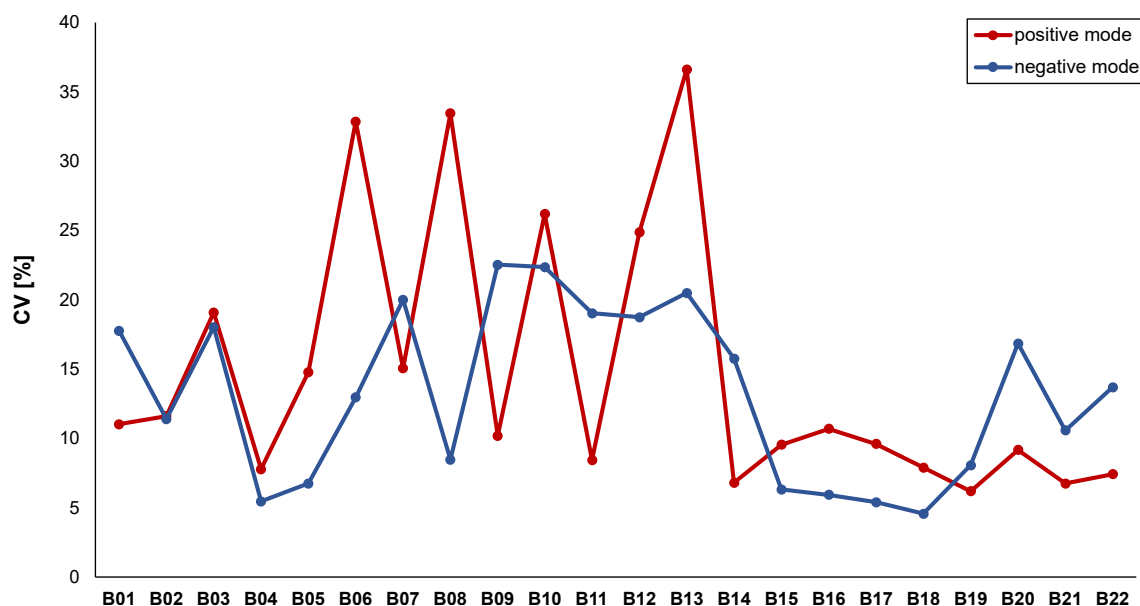


Figure 4.2: Diagram showing the batch-specific coefficient of variation in main study batches. The CV is depicted as the median CV value determined over all internal standards detected in the respective mode and is based on the raw peak intensities in QC samples.

Based on those 15 and 13 IS, respectively, QC charts were prepared in the same manner like for the prestudy cohort (Figure 4.3). The charts reveal stronger batch effects, i.e. inter-batch intensity drifts, than observed in the prestudy data due to the much longer period of data acquisition (one year for main study compared to 3 weeks for prestudy), but still show proper measurement quality over the entire cohort. While the intra-batch intensity level was mostly stable, intensity shifts within a batch were observed for batch 13 in positive ion mode, most probably caused by measurement fluctuations, since clinical and QC samples were affected similarly but sample preparation was performed on different days. Both intra- and inter-batch variations were addressed by post-acquisitional data normalization during the data processing procedure. Samples, being outside the control level, were either remeasured or removed during data processing.



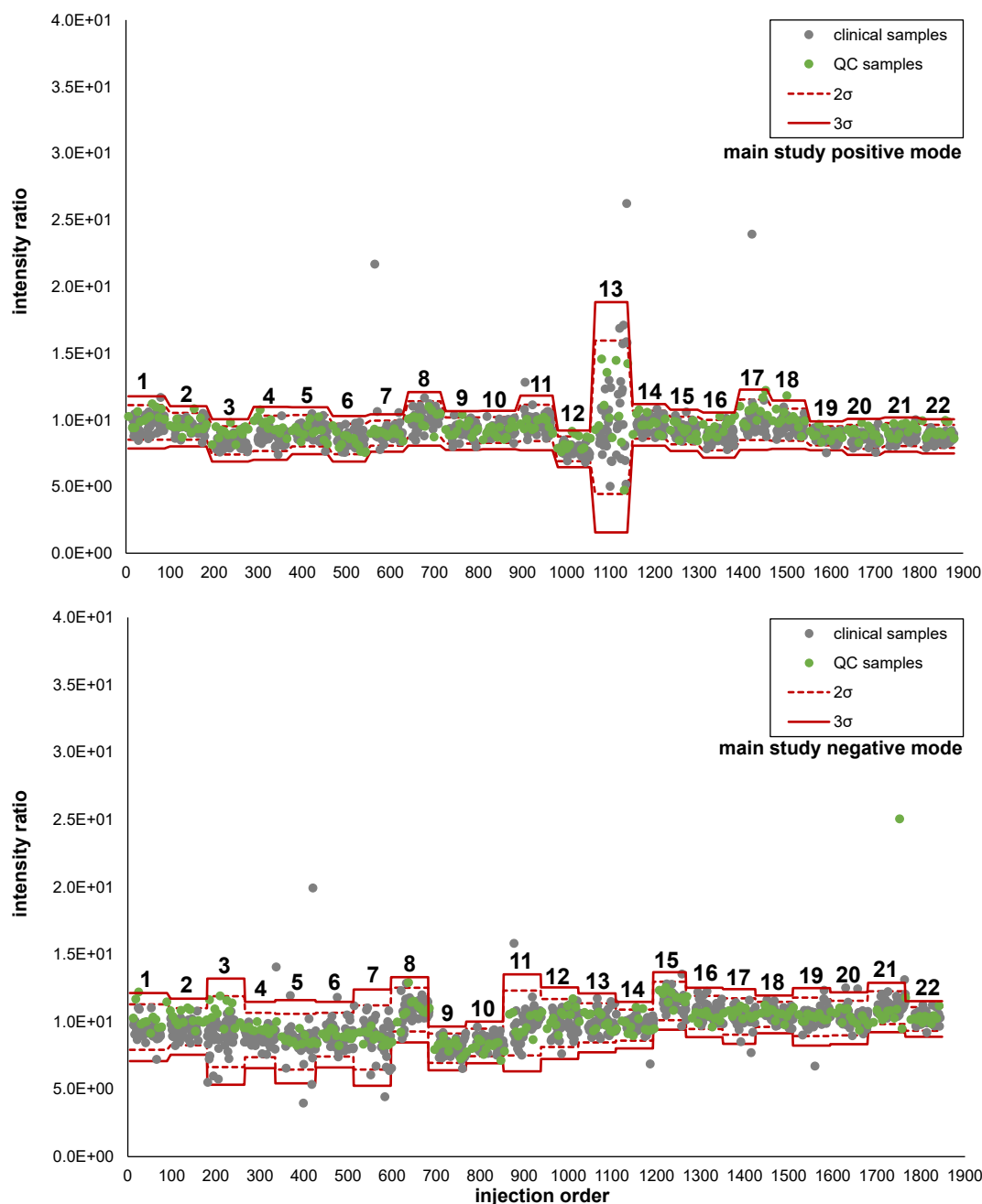


Figure 4.3: Main study quality control charts showing peak intensity ratios of internal standards in clinical and QC samples plotted against the injection order over all 22 acquisition batches in positive and negative ionization mode each. Depicted intensity ratios were obtained by referring the sum peak intensities of all internal standards in the respective ionization mode (pos: 15 IS; neg: 13 IS) to the peak intensity of 18:1(d7) LPC. Batch-wise calculated  $\pm 2\sigma$  and  $\pm 3\sigma$  from the mean indicate warning and control levels, respectively.

In the present datasets, the assessment of measurement quality was based on a quality control lipid extract being prepared in advance for the entire sample cohort. Thus, an identical QC sample, stored in aliquots, was included in all batches. However, in the design of future

large-scale omics studies, the inclusion of a second QC sample matrix could be reasonable. Such a QC sample could be a representative of each batch, i.e. being prepared as a pool of the lipid extracts of the respective batch. Through the use of two QC samples, quality assessment might be more robust and informative, since each batch can be compared and adjusted to other batches by a study QC sample (i.e. identical sample matrix included in all batch sequences of the entire cohort) and sample processing is under control of a batch QC sample (i.e. pool of batch-specific lipid extracts).

### 4.3 Overview of the data processing workflow

The main bioanalytical challenge of the clinical platelet lipidomics study was the development of a data processing workflow to handle these large amounts of data and translate the measurement output into a representative lipidomics datamatrix. While data processing of small sets of samples being acquired in one batch is well established by the use of automated data processing tools like MS-DIAL [85, 117, 130, 191, 201], the same workflow is not similarly transferable to the processing of data acquired in many batches with a total of more than thousand injections including clinical and QC samples. The simultaneous processing of all batches is not applicable due to computing capacity limitations, since SWATH files are relatively large ( $\sim 350$  MB in *ABF*-format) and the majority of processing steps like inter-batch retention time correction, feature alignment, and MS2 spectra deconvolution are computationally intensive tasks. Thus, the established workflow had to be modified and adjusted to be suitable for the application in large-scale lipidomics.

Alternatively to an untargeted data processing strategy, targeted data extraction approaches were developed, firstly reported in 2012 for proteomics by Gillet et al. [126]. Calderon et al. [66] and Cebo et al. [200] identified lipids in untargeted DIA data based on precursor and fragment ions in QC samples of the respective study cohort using an *in-silico* generated target list. The obtained target list was then applied for targeted feature extraction of the entire sample set. However, with this strategy, less features are considered than in an untargeted feature extraction approach (typically a few thousand against 611 [66] or 457 [200] lipid species, respectively) and valuable information might remain unrecognized when restricting data mining to such a target list, e.g. unexpected or unknown lipids of potentially highly interesting biological relevance. With DIAMetAlyzer, DIA data are evaluated using targeted feature extraction based on an assay target library constructed from DDA data [166]. In the present large-scale study, a targeted data extraction strategy is proposed based on an initial step of untargeted data analysis using MS-DIAL, being the basis for the preparation of a reference target list, which is subsequently applied in batch-wise targeted data extraction. The entire developed workflow from data acquisition to the final lipidomics datamatrix is presented in a flowchart in Figure 4.4.

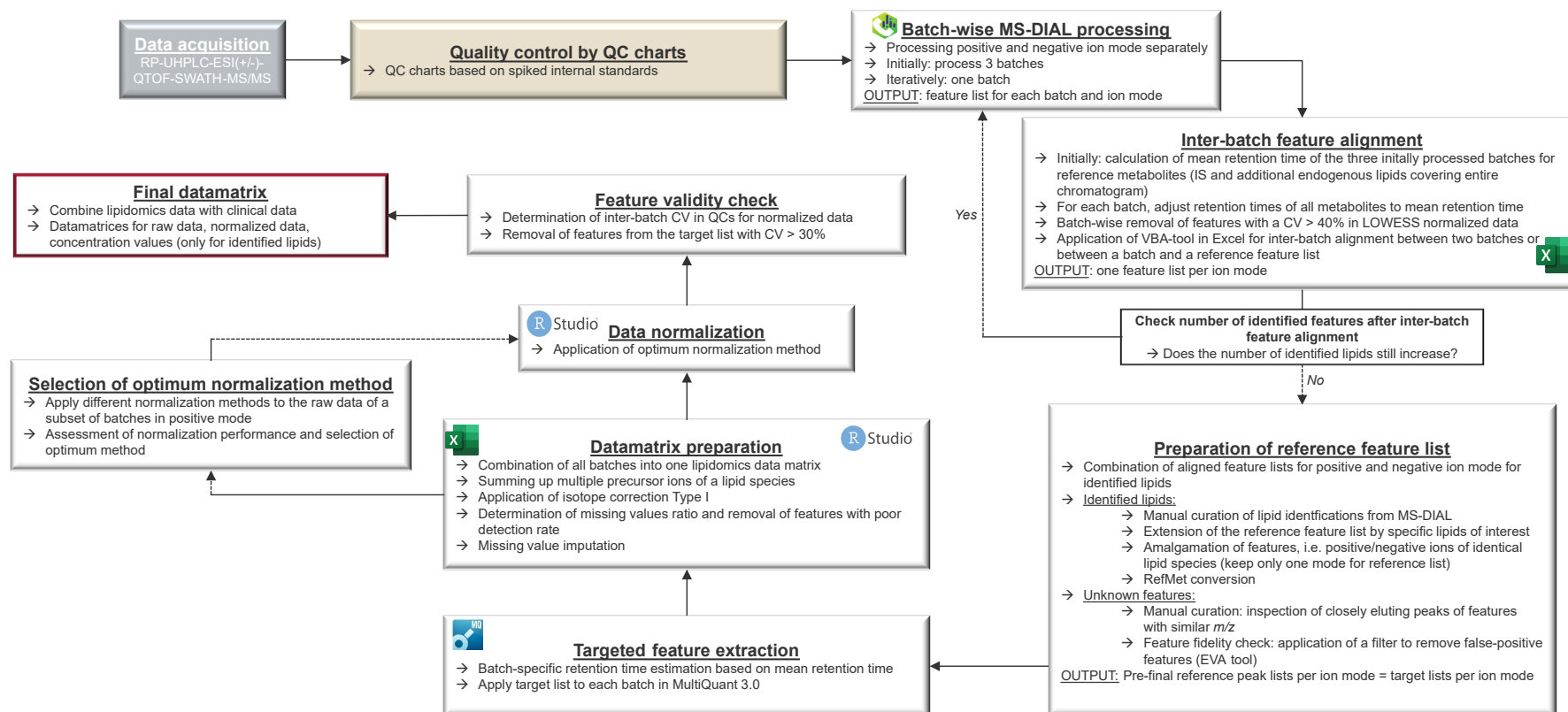


Figure 4.4: Flowchart of the finally developed data processing workflow.

The shown workflow was the final strategy to prepare the lipidomics datamatrix of the main study cohort and remarks the central theme of the results discussion. During the prestudy data analysis, the inter-batch feature alignment tool was developed enabling the alignment of the MS-DIAL processing output from several batches. The design and performance of this tool is discussed in the following sections. Based on the fundamental processing concept from the prestudy data, the procedure was refined during the main study data processing, especially in parts of data curation of the reference target list as well as the datamatrix preparation after targeted data extraction of all batches. Dedicated sections for these processing steps follow the inter-batch feature alignment tool description. Finally in this chapter, the value of global lipid profiling in a clinical cohort is highlighted by several aspects of biological interpretation regarding the role of the platelet lipidome in clinical or therapeutic aspects.

## 4.4 Inter-batch feature alignment

The objective of data processing is the translation of the LC-MS/MS measurement output into a representative feature(pair of  $m/z$  and retention time)-response-matrix. Here, representative means that the corresponding response is available for all listed features in all study samples. Thus, in untargted data analysis, peak detection is firstly performed in each sample, and secondly, detected features are aligned between samples by intra-batch and, in case of multiple batch acquisition, inter-batch comparisons. In the present workflow (Figure 4.4), initial data processing, i.e. from peak detection to intra-batch feature alignment, was performed in MS-DIAL for all (prestudy) or a subset of batches (main study). The selected batches were converted into *ABF*-format and processed with MS-DIAL using the processing parameters shown in Table A.1 for the prestudy and Table A.2 for the main study batches. In MS-DIAL, data were extracted as raw peak heights and LOWESS normalization was performed prior the results feature lists were exported as both raw and normalized peak heights. To merge the batch-wise feature lists from MS-DIAL without reduncancies, a tool for the alignment of features from different datamatrices was developed. The goal of inter-batch feature alignment was the generation of a representative target list for the study cohorts building the base for targeted feature extraction.

### 4.4.1 Tool design and application

The inter-batch feature alignment tool was developed in Visual Basic for Applications (VBA) implemented in Microsoft Excel. In general, features from different batches are aligned based on similarity in accurate mass and retention time. Mass accuracy is a parameter depending on the instrument and is similar for all batches. However, to compensate for shifts in retention time, e.g. due to batch effects, retention time adjustment prior to the alignment is mandatory for a confident alignment result. Therefore, linear correlation based on retention times of internal standards and a small set of manually aligned features covering the entire chromatographic range ( $\sim 2$  features/min) was applied in each ionization mode separately.

Mean retention times of three batches (all in the prestudy cohort; B06, B07, B15 as initially processed batches in the main study) were calculated for those features, and plotted against the batch-specific retention time as shown in Figure 4.5. Using the resulting linear equation, retention times of all features in each batch were adjusted to the mean retention time of the prestudy and main study cohort, respectively. The equations for retention time adjustment of all batches subjected to inter-batch feature alignment are summarized in Table A.5.

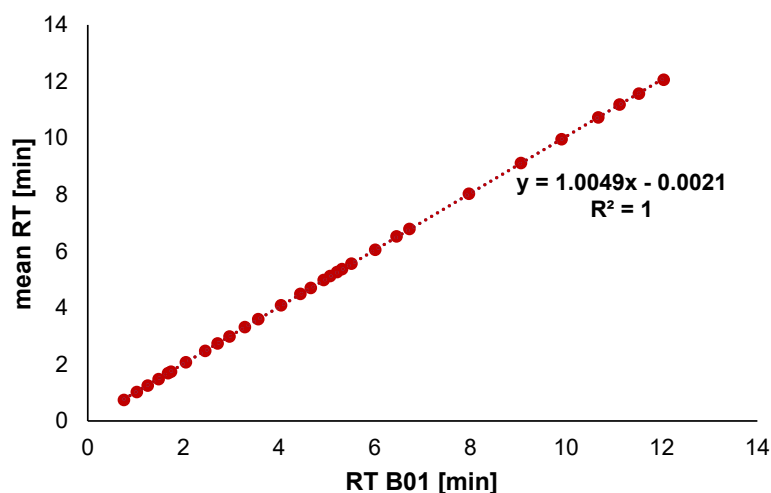


Figure 4.5: Linear retention time correlation of a specific batch against the mean retention time from three batches. Here, the mean retention time from prestudy batches B01, B02, and B03 is plotted against the batch-specific retention time of prestudy B01. Each dot represents a lipid feature.

For inter-batch feature alignment based on feature similarity, limits for mass accuracy ( $\Delta m/z$ ) of precursor masses and retention time ( $\Delta RT$ ) have to be defined depending on instrument performance and inter-batch variance after retention time adjustment. Two feature tables are aligned with each other, either originating from acquisition batches or a reference feature list (alignment result of a former step). The batch or reference feature list with more feature entries is used as the "reference table", to which the table "batch for alignment" is aligned. The following parameters from MS-DIAL processing and retention time adjustment are used to prepare the two tables:  $m/z$ , adjusted retention time, PeakID, metabolite annotation, and the peak intensities for the QC samples acquired in the respective batch. How-to-use instructions are summarized in the first sheet of the Excel template for inter-batch feature alignment (Figure 4.6).

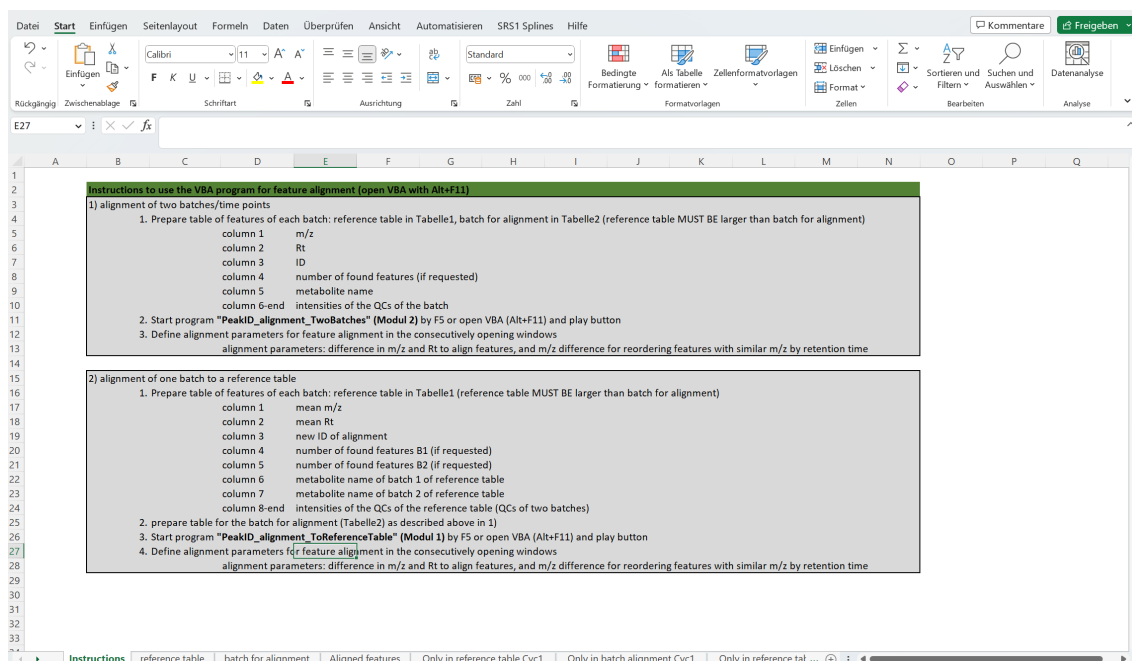


Figure 4.6: Starting Excel sheet with instructions for the application of the VBA alignment tool.

When starting the tool, the user is asked to define the alignment parameters in pop-up windows. In the prestudy, limits of  $\Delta m/z = 0.01$  and  $\Delta RT = 0.1$  min (in main study:  $\Delta RT = 0.16$  min) were used as parameters. During processing, features in each peak table are ordered by increasing  $m/z$  prior to alignment automatically. In addition, features within a very tight  $m/z$ -range (user-defined range called  $\Delta m/z$  closeness, set to 0.016) are rearranged by increasing retention time. By a step-wise comparison of each feature from the “reference table” with the features in the “batch for alignment”, feature alignment is performed.

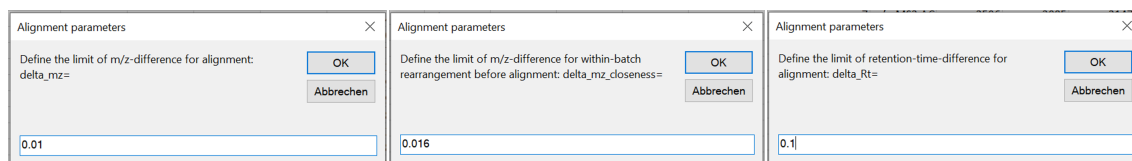


Figure 4.7: Pop-up windows to define alignment parameters.

After the alignment, features are distributed into three tables: a table including aligned features, and two tables (one for each batch) including features for which no alignable counterparts were found. Both latter tables are further used for another run of batch alignment, as due to the step- and pair-wise procedure possible alignments might be missed. In total, three alignment cycles are performed (alignment of 1. complete peak tables from both batches, 2. non-aligned feature tables from both batches after 1st cycle, 3. non-aligned feature tables

from both batches after 2nd cycle) automatically. Finally, the alignment result from three cycles is summarized in three tables, one with aligned features and one table for each batch with non-alignable features.

To assess misalignment of features, a measure was chosen that compares peak intensity of QC samples for a given feature, the zeta value, as recommended by Dunn et al. [134]:

$$\zeta_j = \frac{|\text{median intensity}_{QC,i,j} - \text{median intensity}_{QC,ref,j}|}{\text{median intensity}_{QC,ref,j}} \quad (4.1)$$

In this calculation,  $i$  represents the batch number that is aligned to a reference  $ref$  (either a reference table or another batch) and  $j$  is a counter of features. Instead of peak intensity, peak area can be used. In addition to the zeta value, the difference in accurate mass (in ppm) was calculated between aligned features from different batches to identify misalignments. The calculation of both measures is implemented in the VBA tool and results are summarized in an “alignment evaluation” sheet. A final score indicates if a given alignment passes the acceptance criteria (zeta value < 0.5, difference in accurate mass < 10 ppm) of both measures (final score = 2), only one (final score = 1), or none (final score = 0). Finally, alignment IDs of features with a score of 1 or 0, hence those alignments which need a manual check, are summarized.

The tool is designed to compare pairwise, a batch either to another measurement batch or to a reference feature list. Thus, inter-batch feature alignment in both pre- and main study was performed in a stepwise procedure as shown in Figure 3.2 in Section 3.2.2.3 using all batches processed by MS-DIAL. A reference feature table from the aligned batches was obtained including all detected features from the batches with metabolite names,  $m/z$ -values as mean values, and retention time as mean values.

While values of  $m/z$  and retention time could be easily combined in a mean value, it is much more difficult to merge metabolite names. Basically, after alignment, metabolite names were evaluated for similarity. In case of similar names in all or in the majority of batches in which a distinct feature was detected, or if a feature was only detected in one batch, the metabolite name was adopted. If metabolite names were different, naming of the aligned feature was performed in the decreasing priority as follows: identified > annotated > unknown (identified: identification score from MS-DIAL larger 0.8, annotated: suggested identification but score less than 0.8). Moreover, identified or annotated metabolite names might differ on the same priority level, e.g. due to differences in fatty acyl composition, constitution, or distribution of double bonds. In those cases, one name was chosen randomly (Excel template for metabolite name alignment). Afterwards, identification of all lipids was manually revised considering identification scores, spectral data, and the lipid class specific homologic series of each feature from the contributing batches (see Section 4.5.1).

### 4.4.2 Performance evaluation

The alignment tool was developed during the prestudy data processing, thus the evaluation of the performance was based on the three prestudy batches. Further, feature lists were taken from positive ion mode and reduced to identified lipids, thus, 638 features from Batch 01, 653 features from Batch 02, and 728 features from Batch 03.

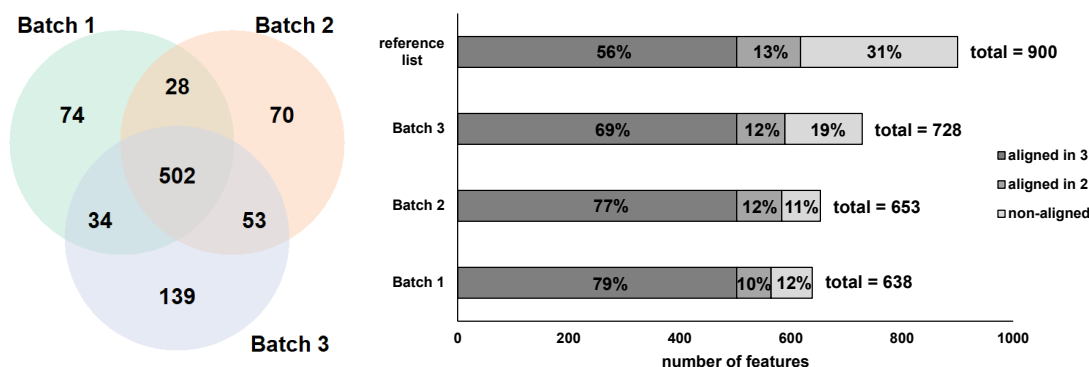


Figure 4.8: Results of inter-batch feature alignment from three prestudy batches. The Venn diagrams visualize the entities of features shared by all three or two batches, or only being present in a single batch. The bar charts show the batch- or reference-list-specific ratios of features shared or being uniquely detected among the three prestudy batches.

As a result of the inter-batch feature alignment of the three prestudy batches, a reference list with 900 features was obtained. From those 900 features, 617 lipids (69%) were aligned between at least two batches (502 features in three batches, 115 lipids in two). The ratio of one-time detected features was 31% ( $n=283$ ). The proportion of features uniquely detected in one batch might be a result of instrumental sensitivity fluctuations since all experimental conditions and processing parameters were identical for all three batches. Especially Batch 03 had significantly more identified features (to Batch 01: 90 more (12%), to Batch 02: 75 more (19%)) compared to the others. Higher sensitivity leads to more detected features when using the same data extraction threshold, and better MS2 spectra for feature identification. However, the large proportion of those features in the reference list supports the benefit of using a representative set of samples from several batches for the generation of a target list for targeted feature extraction instead of only using a single batch.

The correct alignment of identical features between several batches is essential for further statistical analysis. Consequently, it is crucial to identify misaligned features, thus those features that are aligned but not identical. Especially closely eluting features with the same precursor mass, e.g. *sn1/sn2*-isomers like LPC 18:0/0:0 and LPC 0:0/18:0, are prone to misalignment, in particular if a shift in retention time is observed between batches. Therefore, a measure ( $\zeta$ , zeta value) to assess misalignment was selected. The zeta value is based on the intensity in QC samples and relates the difference of median intensities from both aligned



batches to the median intensity of the reference table. A zeta value above a selected level is defined to be a suspect and needing manual curation of the alignment. In addition, the mass accuracy in ppm ( $\text{ppm} \leq 10$ ) of the aligned features is calculated as another performance parameter. By combining both measures, a final score is generated indicating whether both, one, or no criterium is passed, hence suspicious alignments are identified and can be checked manually.

For selecting a limit for the zeta value, several considerations were taken into account. Firstly, choosing the median value for zeta value calculations was favorable over the mean, since the effect of outliers on the result could be avoided. Moreover, the measure has to be sensitive enough to identify misalignments, thus zeta limit should be as low as possible. In contrast, the limit needs to be as high as possible to really only highlight the suspicious alignments that require special attention. In addition, the zeta value limit should be insensitive to inter-batch intensity variations of identical features, i.e. due to batch effects or instrumental shifts, and sensitive to different intensity levels due to misalignments. For the zeta value limit selection, the distribution of zeta values of the inter-batch feature alignment of Batch 01, Batch 02, and Batch 03 is plotted in Figure 4.9.

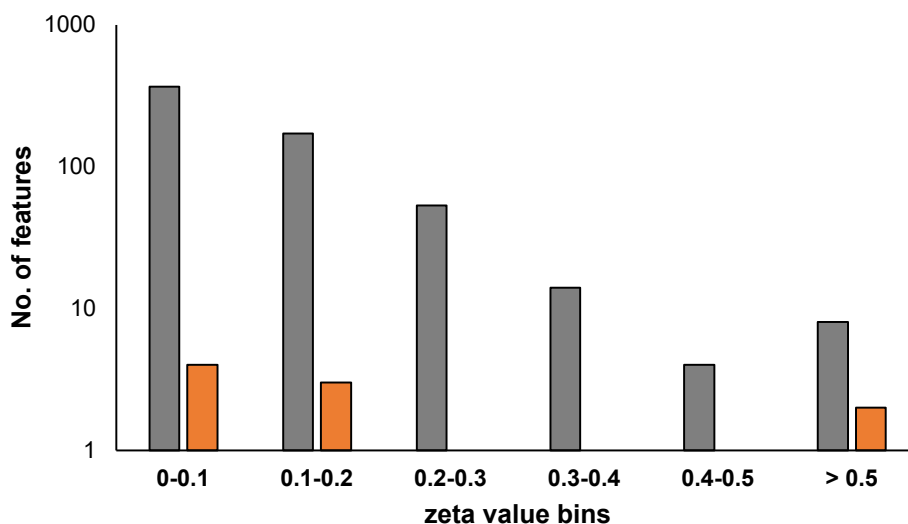


Figure 4.9: Distribution of zeta values over all aligned identified features. Bars in grey refer to all alignments within a zeta value bin, the orange bars indicate misalignments among all alignments in this bin.

The majority of feature alignments had a zeta value of less than 0.5, i.e. 98.7% of alignments of only identified lipids (aligned features: 617). A zeta value of 0.5 means that median intensity of a batch is half-times lower or 1.5-times higher than the median intensity in the reference batch. The mean zeta value of all alignments was 0.11. For 1.3% of the alignments, the zeta value was  $>0.5$ . Feature alignments were manually checked for all zeta values to determine where misalignments occurred. For suspicious alignments, peak spotting maps and the XIC of each feature in each batch were extracted from MS-DIAL and compared for similarity. As a

result, only a very few misalignments occurred which had zeta values in the range of 0-0.1 (4 alignments, 1.1% of this bin), 0.1-0.2 (3 alignments, 1.8% of this bin), and  $>0.5$  (2 alignments, 25% of this bin). Inspecting those alignments revealed that usually two closely eluting, poorly separated peaks were assigned. As an example, Figure 4.10 shows the wrongly aligned isomers of PE 36:5, the second eluting isomer from Batch 01 was matched with the first eluting peak from Batch 02.

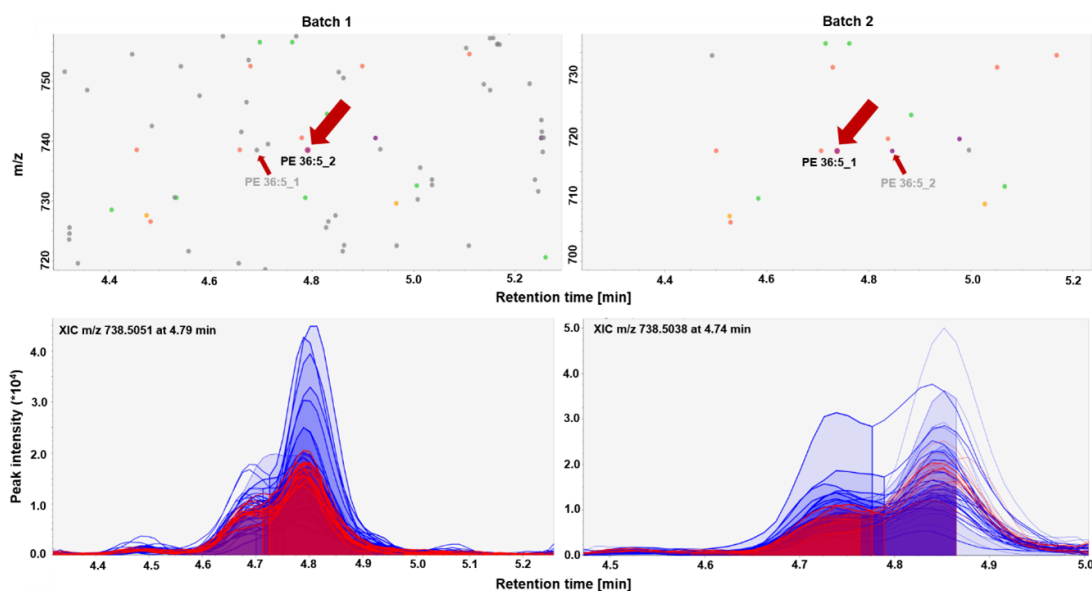


Figure 4.10: Misalignment of PE 36:5 isomers with zeta value of 0.8 in the alignment of Batches 01 and 02. Aligned peak spotting maps (top panel) and respective XICs of PE 36:5 (precursor exact  $m/z$  738.5068  $[M+H]^+$ ) (bottom panel) from MS-DIAL processing were compared between Batch 01 and Batch 02, revealing not identical peaks were used for the alignment. Color codes: QC samples = red, clinical samples = blue.

Here, two closely eluting peaks were mixed up in the feature alignment which was detected by a zeta value of 0.8, hence the alignment was not correct and the aligned features do not represent the same lipid species. However, it has to be noted that these misaligned isomers of PE 36:5 were aligned correctly when the alignment is performed with the entire feature list, i.e. including identified and unknown features. This is caused by the reduction of the peak list to only identified lipids. While two isomers were identified in Batch 02, only one of the two detected isomers was identified in Batch 01, thus the non-identified, as unknown detected feature was excluded from the peak list when only aligning annotated features.

Since the proportion of misalignments was increasing with higher zeta values, a limit of 0.5 was chosen as a good measure to identify misalignments by simultaneously generating a manageable number of feature alignments for manual inspection.

To further assess the alignment performance, all features were manually inspected to determine the proportion false-positive as well as false-negative alignments. True-positive (TP)

alignments are features representing the same peak, while false-positives (FP; =misalignments) do not. False-negatives (FN) are features which were not aligned to a correspondent peak in another batch although an appropriate feature was present, true-negatives (TN) were correctly non-aligned features. The results are shown in a confusion matrix in Table 4.5.

Table 4.5: Confusion matrix showing true-positive and true-negative as well as false-positive and false-negative alignments among all features of Batches 01, 02, and 03; true/false refer to identical/non-identical features in different batches, positive/negative refer to an alignment or non-alignment of features

|              | positive |      | negative |      |
|--------------|----------|------|----------|------|
|              | absolute | [%]  | absolute | [%]  |
| <b>true</b>  | 608      | 98.5 | 281      | 99.3 |
| <b>false</b> | 9        | 1.5  | 2        | 0.7  |
| <b>total</b> | 617      | 100  | 283      | 100  |

The evaluation revealed a very good performance of the inter-batch alignment tool in terms of correctly aligning equal features from different batches with a low rate of wrongly aligned features or missed alignments. Among all alignments ( $n=617$ ), a precision ( $=TP/(TP+FP)$ ) of 98.5% was achieved remarking the positive prediction value. The recall value ( $=TP/(TP+FN)$ ) was 99.8%, since only two features were wrongly not aligned although being present in two batches, thus one alignment was missed. However, 99.3% of the non-aligned features ( $n=281$ , combined from three batches) were truly only detected in one of the three batches. Both precision and recall values are typical parameters to assess the accuracy of an algorithm [202] and here indicate a reliable tool performance of the inter-batch feature alignment.

#### 4.4.3 Effect of alignment parameters on performance

Prior to performing the inter-batch feature alignment, three alignment parameters have to be defined: mass accuracy range ( $\Delta m/z$ ) within feature precursor ion masses are defined to be similar, retention time range ( $\Delta RT$ ) specifying the elution window of identical features, and a parameter called  $m/z$  closeness ( $\Delta m/z$  closeness). This parameter determines a range of precursor ion masses within features, primarily ordered by increasing precursor ion masses, are reordered by increasing retention time. Three different levels per parameter were tested to evaluate their influence on the alignment result.

Initially, a medium level of parameters was defined ( $\Delta m/z$ : 0.01 Da,  $\Delta RT$ : 0.1,  $\Delta m/z$  closeness: 0.016 Da (1.6-times  $\Delta m/z$ )). These medium parameter levels were also used for all above discussed evaluations. Other tested parameter combinations are summarized in Table 4.6. To assess the effect of each parameter individually, two parameters were kept constantly at medium level, while the third one was changed. With each set of parameters, inter-batch feature alignment was performed for all identified features of Batches 01 and 02, and was evaluated for the number of aligned features in each approach (Figure 4.11).

Table 4.6: Parameter combinations to evaluate the effect of parameter levels on alignment result

| parameter levels |      | $\Delta m/z$ | $\Delta m/z$ closeness | $\Delta RT$ |
|------------------|------|--------------|------------------------|-------------|
| medium           |      | 0.010        | 0.016                  | 0.10        |
| $m/z$ range      | low  | 0.005        | 0.016                  | 0.10        |
|                  | high | 0.020        | 0.016                  | 0.10        |
| $m/z$ closeness  | low  | 0.010        | 0.010                  | 0.10        |
|                  | high | 0.010        | 0.020                  | 0.10        |
| RT range         | low  | 0.010        | 0.016                  | 0.05        |
|                  | high | 0.010        | 0.016                  | 0.20        |

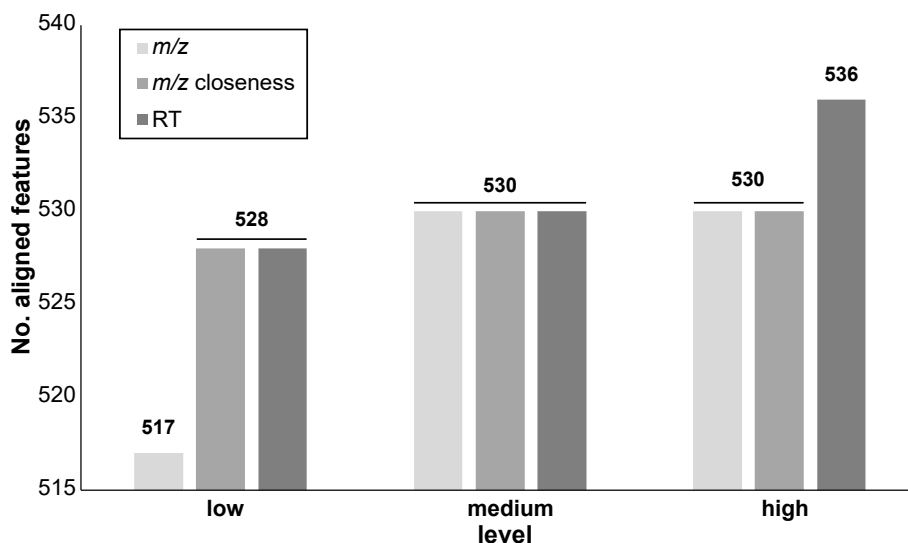


Figure 4.11: Bar chart plotting number of aligned features obtained by inter-batch feature alignment of Batches 01 and 02 with applying 3 different levels of each alignment parameter. Medium level was  $\Delta m/z$ : 0.01 Da,  $\Delta RT$ : 0.1 min,  $\Delta m/z$  closeness: 0.016 Da, testing each parameter was performed by varying levels of one parameter while keeping the others at medium level.

Irrespective of the parameter, each test of low-level values resulted in lower number of aligned features than it was obtained for the medium level of the parameters ( $n=530$ ). For low-level  $\Delta RT$  and  $\Delta m/z$  closeness, two alignments less were observed ( $n=528$ ). The lower number is caused by missing one correctly aligned and one misaligned feature. For  $\Delta m/z$  only 517 features could be aligned between Batch 01 and Batch 02. It was observed that identical features were not aligned due to too narrow mass accuracy limits. However, a wider range of mass accuracy (i.e. high level) did not cause an increase in aligned features compared to medium parameter level. The same effect was observed for the parameter  $\Delta m/z$  closeness.

For the retention time range, a consecutive increase in aligned features was obtained with increasing  $\Delta RT$  limits. In total, six additional alignments resulted from inter-batch feature alignment with a high level for retention time range. Manually checking the additionally

aligned features revealed that all were misalignments, thus the higher number of alignments did not go along with an improvement. As a conclusion, the medium level of all parameters ( $\Delta m/z$ : 0.01 Da,  $\Delta RT$ : 0.1,  $\Delta m/z$  closeness: 0.016 Da) performed best in terms of number of correct alignments. The parameter  $\Delta m/z$  turned out to be the most critical one in terms of underestimation of alignable features, while the parameter  $\Delta m/z$  closeness had only a minor effect. In contrast,  $\Delta RT$  was the most crucial parameter in terms of overestimating the number of alignments.

#### 4.4.4 Improvement in lipid identification

A general beneficial effect of considering a large number of samples for the generation of a reference target list is the enhancement of feature identification. The more samples are aligned, the higher is the possibility of obtaining a MS/MS spectrum with sufficient data quality for identification. Especially in DIA approaches like SWATH, the quality of MS/MS spectra is crucial due to interferences with cofragmented coeluting compounds [131, 203] and a larger number of samples might positively contribute to the identification rate. Therefore, it was evaluated how the inter-batch feature alignment contributed to feature identification in the alignment of Batches 01, 02, and 03 considering all detected features. The results of this evaluation are summarized in Figure 4.12.

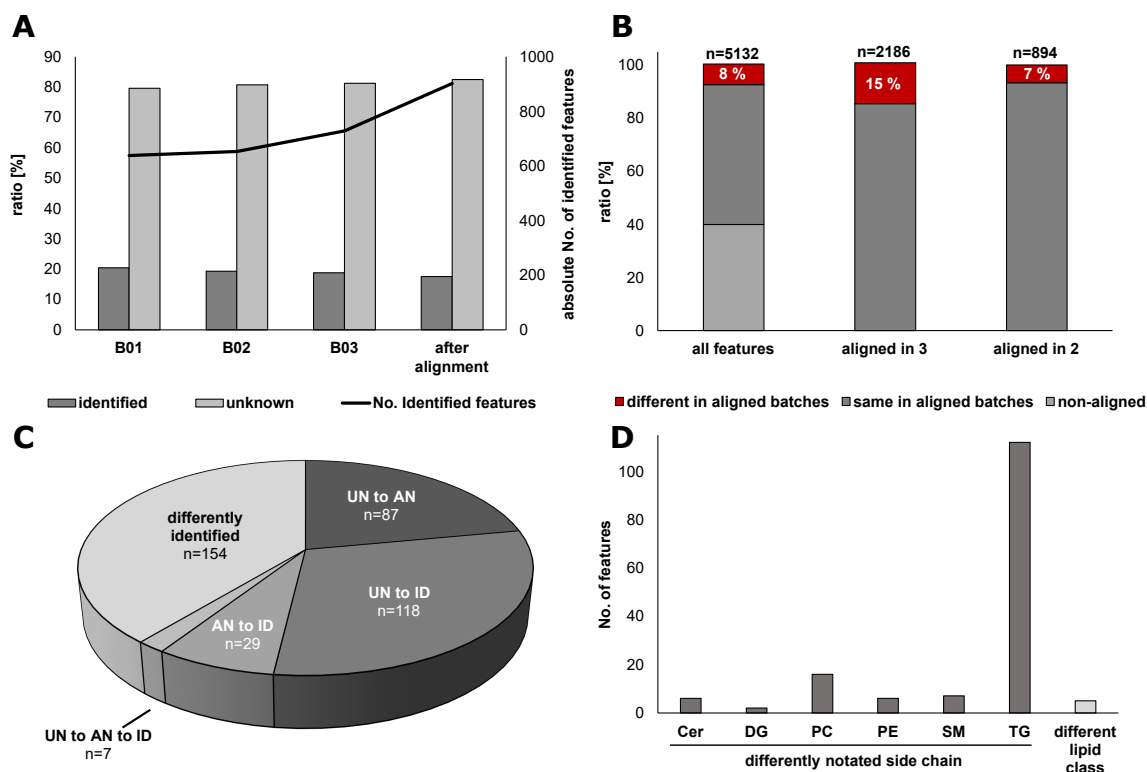


Figure 4.12: Beneficial effect of inter-batch feature alignment on lipid identification. (A) Bar chart displaying the ratio of identified features to all detected in each batch and in reference feature list after alignment as well as the absolute number of annotated features. (B) shows the ratio of features having different identification levels and/or metabolite names for the same feature in the contributing batches. (C) Pie chart showing the types of improvement and its distribution of 395 aligned features differing in identification level or metabolite name (UN=unknown, AN=annotated, ID=identified). (D) Distribution of lipid classes which were differently identified between batches.

In a first step, the ratio of identified features compared to the total number of detected features was evaluated for each of the three batches as well as in the reference feature list after alignment (Figure 4.12A). A proportion of around 20% of the detected features was identified in all batches by MS-DIAL data processing. While no improvement in identification rate was observed, a gain of at least 24% (based on the number of identified features from Batch 03) was obtained on the total number of identified features (638 (Batch 01), 653 (Batch 02), and 728 (Batch 03) to 902 (after alignment)). Hence, by increasing the number of batches and therefore samples contributing to a reference feature list, the number of successful lipid identifications was significantly increased as compared to an approach considering only a low number of samples.

Moreover, lipid identification was evaluated in more detail by comparing metabolite names between the three batches. Based on the identification procedure from MS-DIAL [117], three “levels” of identification were assigned, 1. Identified (ID), 2. Annotated (AN), 3. Unknown

(UN). Identified lipids have an identification score larger than 0.8 in MS-DIAL (a score of 1 means perfect match with the reference in the database). For annotated features, a suggestion for an identification is made, but with a score less than 0.8 (insufficient quality of MS/MS spectra). Identifications might be similar, different, or improved in the contributing batches. Improvement was defined as achieving a higher level of identification in a priority order of Identified > Annotated > Unknown. The comparison revealed that after alignment, 8% of the aligned features (395 out of 5132 features) were different in terms of identification level or metabolite name (Figure 4.12B). Within 2186 aligned features being detected in all three batches, it was 15% and for 894 features being aligned from two batches it was 7%. From the proportion of 8% (n=395 features, Figure 4.12C), 87 former unknown features were annotated after alignment (UN to AN). For another 118 Unknowns, a lipid identification was obtained (UN to ID). Moreover, 7 features were unknown, annotated, and identified over the three aligned batches. Additionally, 29 aligned features improved from annotated to identified. Hence, 60% of these 395 features were improved in terms of identification level. Besides those, 154 features differed in the suggested metabolite name between batches. Lipid identifications were either different due to differently annotated side chain compositions, which occurred in glycerolipids (DG, TG), glycerophospholipids (PC, PE), and sphingolipids (Cer, SM), or due to different lipid class assignment (e.g. PC and LPC species; Figure 4.12D). Different identification was a result of non-sufficient spectral data on side chain compositions in at least one contributing batch. In conclusion, the inter-batch feature alignment both improved identification and highlighted features with special need for manual curation of lipid identification. After manual curation, lipids were annotated as recommended by Liebisch et al. [46].

The inter-batch feature alignment of the three prestudy batches revealed a significant increase in identified lipids. However, due to the limited number of batches in this cohort, all acquired batches were aligned, thus all measured samples were used for the generation of a reference target list. Considering all batches is not a suitable strategy in the main study cohort since 22 batches in both polarity modes were acquired. Consequently, a subset of batches being representative for the entire sample cohort was used for the reference target list generation by inter-batch alignment. Initially, three batches (Batch 06, Batch 07, Batch 15) were processed by MS-DIAL and subsequently aligned similarly to the prestudy processing workflow. Afterwards, additional and randomly selected batches are processed and aligned iteratively until the number of identified features was not further increasing in the obtained reference feature list (see workflow in Figure 4.4). Finally, eight batches (Batches 06, 07, 11, 15, 18, 19, 20, 21) contributed to the reference feature list, hence representing 35% of the sample cohort (=367 clinical samples). The development of the ratio of identified lipids among all detected features as well as the absolute values in all contributing batches and the obtained reference lists is shown in Figure 4.13.

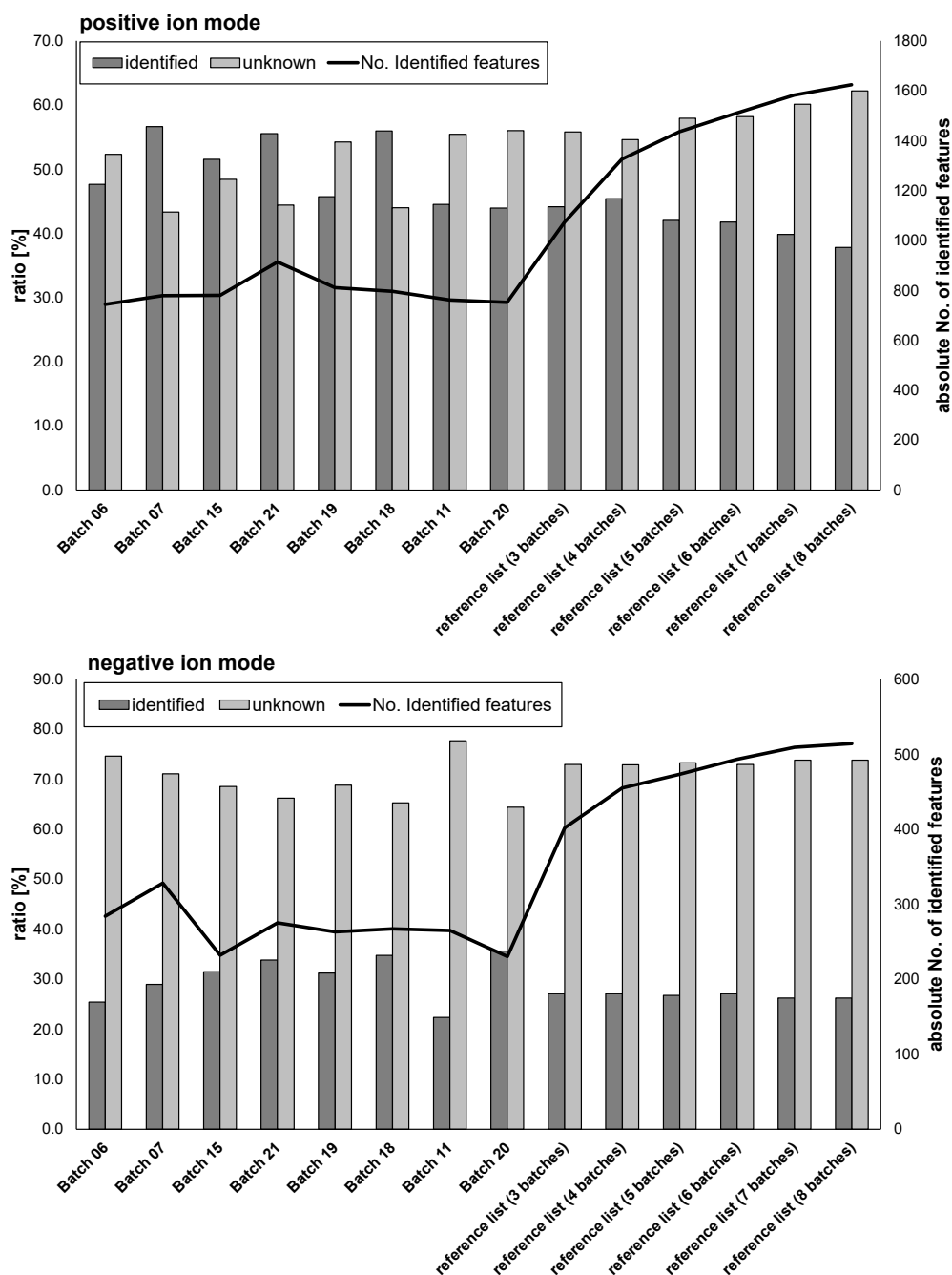


Figure 4.13: Development of proportion and number of identified lipids among all detected features in main study data by inter-batch feature alignment. In the upper diagram, the data in positive ion mode are depicted and in the lower one, the data from negative ion mode. Ratios of identified features (bars in dark grey) as well as unknown features (bars in light grey) and the absolute number of identified lipids (line) are shown for each contributing batch and all created reference lists.

Similar to the prestudy results, the proportion of identified lipids among all detected features was not improved by the alignment of several batches, being on the same level (at 27% in



negative ion mode) or slightly decreasing (to finally 38% in positive ion mode). In general, the ratio of identified features was lower in reference feature lists than in individual batches. This might be caused by randomly occurring unknown features, i.e. artefacts, being detected in any of the batches. Since no data curation was performed up to this workflow step (except of blank subtraction during MS-DIAL processing), those features are still included but are aimed to be removed during the following curation steps. In addition, it has to be noted, that the lipid identifications are the non-reviewed annotations from MS-DIAL processing, hence, manual curation of the identifications is lacking here as well.

Both bar charts indicate the significant increase in the number of identified features through the alignment of feature lists from several batches. Compared to a single batch, a gain of 178% in positive ion mode (1625 to 914 features in "best" batch) and 157% in negative ion mode (514 to 328 features in "best" batch) was achieved. Moreover, by the step-wise addition of batches, the number of identified features evolved from 1074 in the reference list of three to 1625 in the reference list of eight batches (+51%) in the positive ion mode and from 402 to 514 features (+28%) in negative ion mode. Furthermore, the increase shows an asymptotic progression with no significant gain between seven and eight aligned batches, thus, the feature list for further targeted data extraction was based on eight batches representing the clinical sample cohort.

#### 4.5 Reference feature list preparation

Through MS-DIAL processing and subsequent inter-batch feature alignment of three and eight batches from the prestudy and main study cohorts, respectively, representative feature lists for both polarity modes were obtained. A summary showing all detected, and among them unknown and identified features, is listed in Table 4.7. Here it has to be noted, that the number of features in positive ion mode of the prestudy data is lower than described in Section 4.4.2 (only data from positive mode were shown) due to the application of a peak count filter on the aligned prestudy reference lists. Based on a peak count filter, rarely detected features can be removed during data processing, i.e. a given feature has at least to be found in a defined proportion of samples among a sample group within batch. Since the sample groups were not confidently known when starting the data processing, the peak count filter in MS-DIAL was kept low (=2%). When the sample groups were confirmed, a peak count filter of 5% was applied after batch alignment in Excel, calculated on the smallest sample group (n=7). Thus, the number of reference list features was reduced from 5828 to 3195 in positive and 1483 to 966 in negative ion mode.

Table 4.7: Summary of reference feature lists after MS-DIAL data processing and inter-batch feature alignment in both positive and negative ion mode for prestudy and main study data

| cohort     | features          | positive mode |           | negative mode |           | combined |           |
|------------|-------------------|---------------|-----------|---------------|-----------|----------|-----------|
|            |                   | number        | ratio [%] | number        | ratio [%] | number   | ratio [%] |
| prestudy   | <b>total</b>      | 3195          | 100       | 966           | 100       | 4161     | 100       |
|            | <b>identified</b> | 768           | 24        | 333           | 34        | 1101     | 26        |
|            | <b>unknown</b>    | 2427          | 76        | 633           | 66        | 3060     | 74        |
| main study | <b>total</b>      | 4299          | 100       | 1962          | 100       | 6261     | 100       |
|            | <b>identified</b> | 1625          | 38        | 514           | 26        | 2139     | 34        |
|            | <b>unknown</b>    | 2674          | 62        | 1448          | 74        | 4122     | 66        |

Considering both polarity modes, 4161 features were included in the reference feature list of the prestudy cohort, among which 26% were annotated as lipids. In the main study data, 6261 features, i.e.  $\sim 2000$  features more, were listed with an identification rate of 34%. The larger number of features is caused by the inclusion of eight in contrast to three batches during the reference list preparation. Furthermore, due to changes in the blood sampling conditions (see Section 3.2.1.1), differences in the pellet platelet count for lipid extraction (prestudy:  $3 \times 10^8$  platelets per pellet; main study:  $2 \times 10^8$  platelets per pellet), and thus adjusted processing parameters in MS-DIAL, i.e. peak detection threshold, differences between both feature lists were expected. The higher rate of annotated lipids (34% to 26%) could be expected as well, since more samples were contributing to the main study reference feature list increasing the possibility of good quality MS/MS spectra being sufficient for lipid identification. In addition, MS-DIAL processing of the prestudy and main study data was two to three years apart with advancements in LipidBlast database in the meanwhile. Consequently, due to the described differences, the data processing workflow for the main study was performed individually and not as an extension of the prestudy processing. However, the list of identified lipids from the prestudy cohort was aligned against the reference feature list from the main study to support the lipid identification in curation of identified and unknown features.

The reference feature lists after MS-DIAL processing and inter-batch feature alignment build the basis for the target list being used for targeted feature extraction. However, data curation of the reference features, identified as well as unknown features, is mandatory to ensure a confident results reporting and improve the study outcome. Hence, the following sections are dedicated to the data curation of identified lipids and unknown features resulting in the final target list for data extraction.

#### 4.5.1 Data curation of identified lipids

Lipid identification is an essential step in untargeted lipidomics data processing, in both studies being performed in MS-DIAL in an automated manner. Basically, detected features are compared to a reference database, i.e. LipidBlast, in terms of accurate mass, isotopic pattern, spectral similarity on MS/MS level, and, if possible, retention time information [117,

170]. The comparison results are expressed in a score ranging from 0 to 1 indicating no to perfect matching of experimental and reference data. However, for automated lipid annotation using *in-silico* libraries often an overestimation of annotated lipids is observed, resulting in misannotated, false-positive lipids [170, 204, 205]. Thus, manual identification validation was performed to confirm lipid identifications and to remove or correct misannotations.

The curation of lipid annotations was based on two major parameters, the MS-DIAL identification score and the retention time. In addition, information from both polarity modes or different adduct types, if available, were combined to support the feature identification. It has to be noted, that the retention time was already used for scoring in MS-DIAL during the prestudy processing, while this was not possible for the main study due to retention time changes in the LipidBlast database. For assigning an identification as valid, the MS-DIAL identification score has to be  $>0.8$  and a feature has to match the lipid class specific homologic series. In reversed-phase separations, the retention time of lipid species within a lipid class is increasing linearly with the precursor ion mass of species with increasing carbon chain length but same degree of unsaturation, or is decreasing with the precursor ion mass of species with the same carbon chain length but increasing number of double bonds. As an example, in Figure 4.14 the homologic series of LPCs in positive ion mode of the main study data is depicted demonstrating the valuable contribution of retention time information on lipid identification.

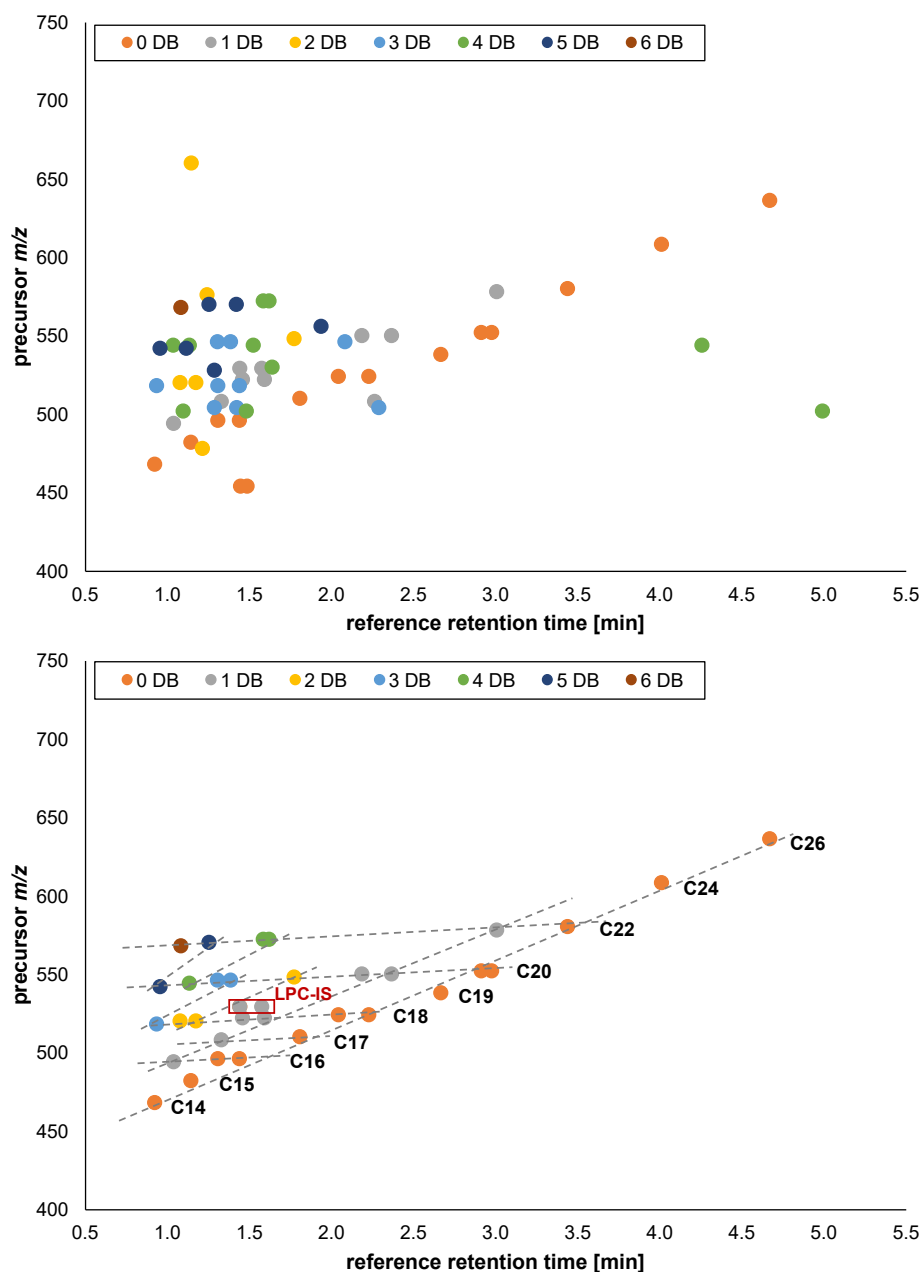


Figure 4.14: Homologic series of LPC species in positive ion mode of the main study data. The precursor  $m/z$  is plotted against the reference retention time. The upper panel shows all detected LPC species after inter-batch feature alignment, the lower panel only represents valid LPC species after manual curation.

In the upper diagram, all detected LPC species after inter-batch feature alignment are plotted. The homologic series is highly indicative for misannotations of several species. After the removal of non-matching LPC species and in accordance with the identification results from MS-DIAL, the lower panel was obtained, representing the set of LPCs included in the reference target list. Here, 60% of the initially detected LPC species were confirmed in their lipid annotation.

The homologous series were prepared for all detected lipid classes with individual charts for positive and negative ion mode as well as different adduct types in pre- and main study data separately, resulting in more than 50 plots. Afterwards, all identified lipids were inspected in MS-DIAL for the identification score and lipid notation level. Firstly, a feature was selected in the peak spotting map in the MS-DIAL GUI enabling the inspection of the respective XIC (Figure 4.15).

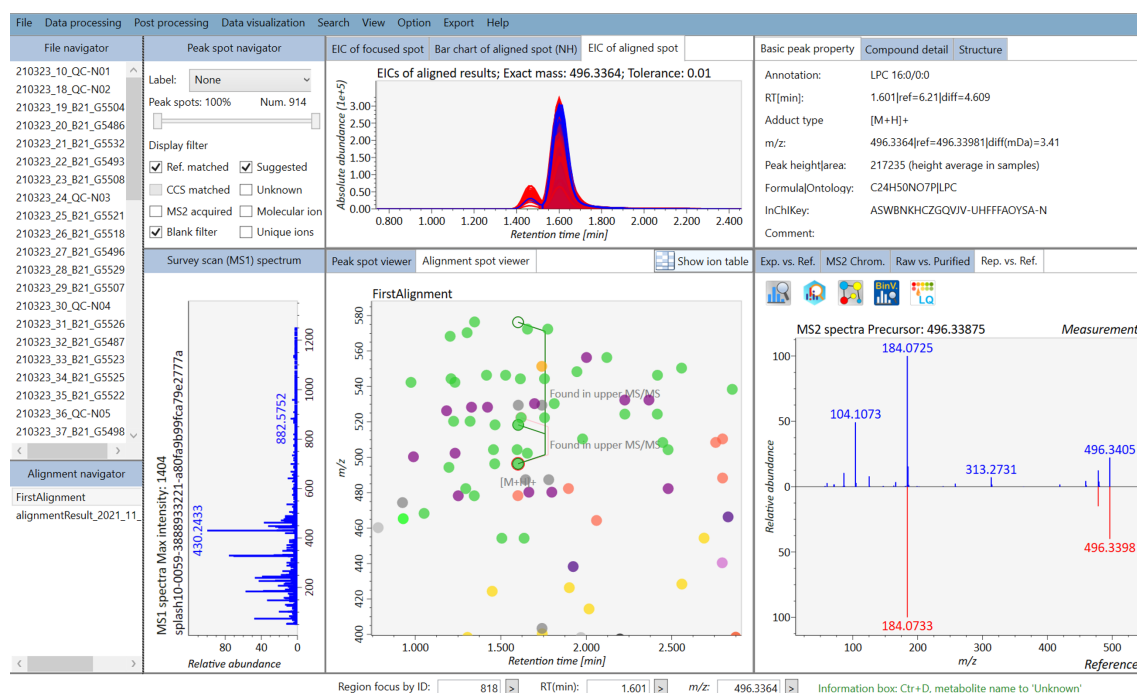


Figure 4.15: Manual curation of automated lipid identification: The GUI of MS-DIAL is shown with the alignment result from Batch 21 in positive ion mode. The alignment spot of LPC 16:0/0:0 was selected (diagram in the lower middle), resulting in the creation of the respective XIC above. On the right top, feature properties are summarized and on the right bottom side, the deconvoluted experimental MS2 spectrum is represented against the reference spectrum.

Based on the spotting map and the XIC, the peak shape and validity of a detected feature was evaluated. Sometimes, even features with an irregular, non-reproducible peak shape were annotated as a lipid. Those features were discarded. The identification performance, in Figure 4.15 already indicated by comparison of the deconvoluted experimental MS2 spectrum with the respective reference spectrum, was further evaluated in the "Compound search" panel (Figure 4.16).

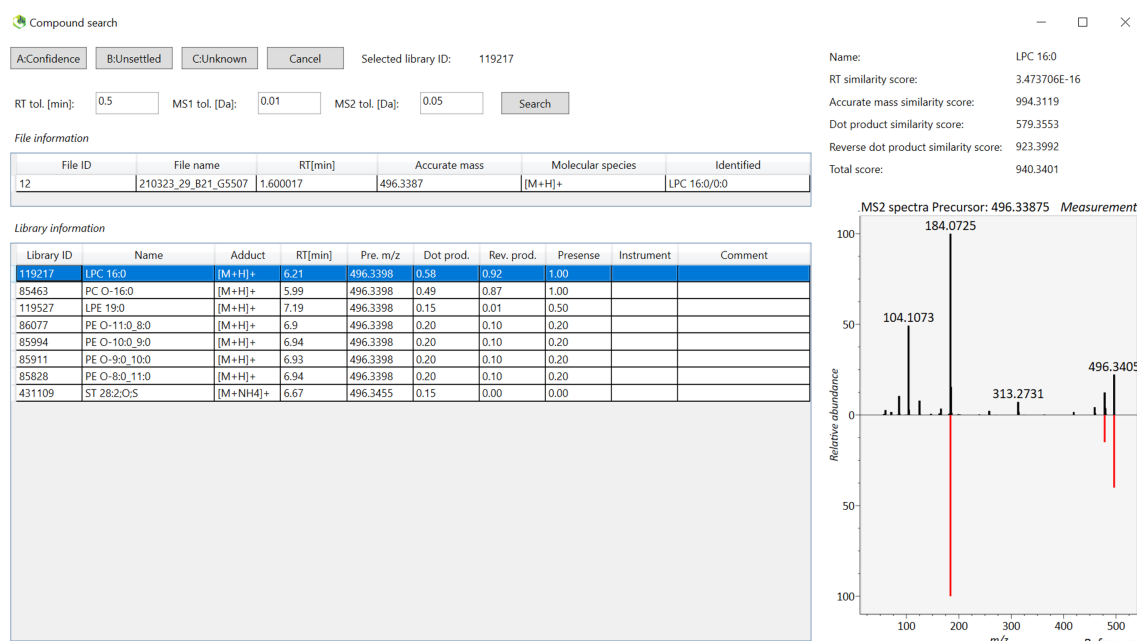


Figure 4.16: Manual curation of automated lipid identification: The "Compound search" panel in MS-DIAL allows the inspection of the total identification score as well as the contributing scores (on the right top and table on the left side). In addition, the deconvoluted experimental spectrum is plotted against the reference MS2 spectrum.

The "Compound search" panel is designed for the curation of the automated lipid identification and provides information regarding the total identification score, the contributing scores (i.e. accurate mass, dot product, reverse dot product, retention time (note: RT was not included in scoring in the main study processing)), and the spectral comparison. A table with potentially matching lipids and the respective identification scores are listed here as well. The spectral similarity is expressed by the dot and reverse dot products. The dot product represents how well an experimental MS2 spectrum matches the reference spectrum, while the reverse dot product score indicates the similarity of a reference spectrum with an experimental one. Consequently, the reverse dot product score is usually higher than the dot product score which is affected by spectral noise. Furthermore, the presence of all potential fragments is given in a score. For the curation of the suggested lipid annotations, all available information were taken together, among which the fragment presence and the reversed dot product turned out to be the most indicative parameters for spectral similarity. In addition, for lipids ionizing in both positive and negative ion mode (e.g. (L)PCs, (L)PEs) or creating several adducts (e.g. (L)PCs, TG), information from these different sources as well as from different batches, were merged to validate lipid identifications.

Based on this strategy, lipid annotations in both prestudy and main study data were curated. However, the curation output was diverse. Besides the confirmation of a suggested lipid annotation or the identification of a misannotated feature, the level of identification could

require some adjustments. For example, if a lipid was annotated with assigning the side chains, e.g. PC 20:0\_22:4, and the similarity in accurate mass and retention time as well as the lipid class specific fragment, here  $m/z$  184.0733, support the annotation as a PC 42:4, but the side chain composition was not confidently confirmed (Figure 4.17), the lipid notation level was reduced.

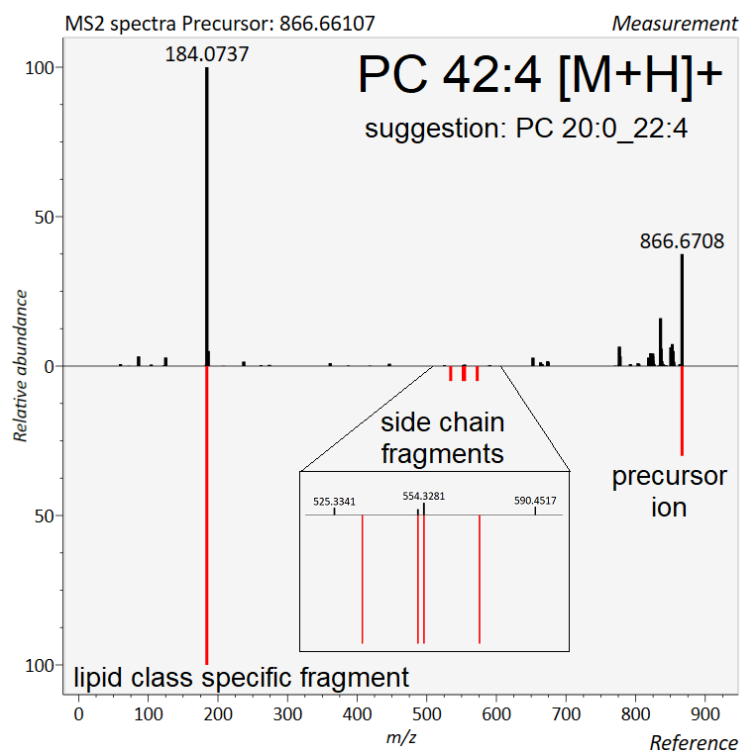


Figure 4.17: Identification curation resulted in annotation adjustment by reducing identification level. The comparison of the experimental and reference MS2 spectra is shown for PC 42:4 [M+H]<sup>+</sup>.

As a conclusion, each identified feature was evaluated with four different potential curation results. Valid lipid annotations were kept as suggested by MS-DIAL (=True Level 1) or the annotation was adjusted due to insufficient spectral information for the suggested annotation level, i.e. adjust to lipid sum notation (=True Level 2). Misannotations were either renamed as "Unknown" (=False Level 1) or removed in case of peaks being false-positive features (=False Level 2). False-positive features might be artefact peaks or a misalignment. Furthermore, in the main study data, redundant lipids were removed, thus applied to features being detected in both polarity modes (=False Level 2). The respective features were kept in the target list in the preferable ion mode, i.e. higher sensitivity. Consequently, the manual curation of lipid annotations led to a significant decrease in identified features in both prestudy and main study datasets. In Tables 4.8 and 4.9, a summary is given for the pre- and main study data, respectively. Since redundantly detected lipids were removed in the main study data, one

table is shown representing lipid identifications from both positive and negative ion mode, while two tables are depicted for the prestudy curation results.

Table 4.8: Data curation of identified lipids in the prestudy dataset presented individually for positive and negative ion mode

| <i>positive ion mode</i>        |                                                            |        |           |
|---------------------------------|------------------------------------------------------------|--------|-----------|
| curation result                 |                                                            | number | ratio [%] |
| <b>True Level 1</b>             | Confirmed lipid annotation from MS-DIAL <sup>1</sup>       | 323    | 42        |
| <b>True Level 2</b>             | Identification level adjustment                            | 352    | 46        |
| <b>False Level 1</b>            | No confirmation of identification, kept as unknown feature | 57     | 7         |
| <b>False Level 2</b>            | Feature removal due to false-positive                      | 36     | 5         |
| <b>total annotated features</b> | prior curation                                             | 768    | 100       |
|                                 | after curation                                             | 675    | 88        |

| <i>negative ion mode</i>        |                                                            |        |           |
|---------------------------------|------------------------------------------------------------|--------|-----------|
| curation result                 |                                                            | number | ratio [%] |
| <b>True Level 1</b>             | Confirmed lipid annotation from MS-DIAL <sup>1</sup>       | 218    | 65        |
| <b>True Level 2</b>             | Identification level adjustment                            | 64     | 19        |
| <b>False Level 1</b>            | No confirmation of identification, kept as unknown feature | 39     | 12        |
| <b>False Level 2</b>            | Feature removal due to false-positive                      | 12     | 4         |
| <b>total annotated features</b> | prior curation                                             | 333    | 100       |
|                                 | after curation                                             | 282    | 85        |

<sup>1</sup> Including internal standards (positive ion mode: n=15, negative ion mode: n=11).

The curation of identified lipids in the prestudy dataset revealed a good annotation performance, 88% of the lipid identification were kept for the target list in positive and 85% in the negative ion mode. While a small proportion of features was completely discarded since being false-positives (positive ion mode: 5%, negative ion mode: 4%), features were reassigned as "Unknown" for 7% and 12% of the annotations, respectively. In total, 675 annotated lipids were kept for a feature target list with 323 features with confirmed lipid identification (42%) and 352 features after identification adjustment (46%) in the positive ion mode. In negative ion mode, 333 features were kept among which 218 features were confirmed in their suggested annotation (65%) and 64 were adjusted in their annotation level (19%). Since the retention time could be used for scoring during automated feature identification, the suggested lipid annotation of many features was correct. However, especially in the positive ion mode, a significant proportion of features was overestimated in terms of the spectral information and the resulting identification level. During automated feature identification in MS-DIAL, the most probable annotation was selected unless the spectral information regarding the side chain composition was insufficient. Thus, these results demonstrate the necessity of manual curation of automated lipid annotation in untargeted omics approaches.

Therefore, the curation procedure was applied for the main study data similarly and the results are shown in Table 4.9.



Table 4.9: Data curation of identified lipids in the main study dataset combined for positive and negative ion mode

| curation result                                                                 | number         | ratio [%] |
|---------------------------------------------------------------------------------|----------------|-----------|
| <b>True Level 1</b> Confirmed lipid annotation from MS-DIAL <sup>1</sup>        | 705            | 33        |
| <b>True Level 2</b> Identification level adjustment                             | 163            | 8         |
| <b>False Level 1</b> No confirmation of identification, kept as unknown feature | 947            | 44        |
| <b>False Level 2</b> Feature removal due to false-positive or redundancy        | 324            | 15        |
| <b>total annotated features</b>                                                 | prior curation | 2139      |
|                                                                                 | after curation | 868       |

<sup>1</sup> Including internal standards (n=41) in both detected modes and considering all detected ion adducts.

The data curation result of the main study data revealed significant differences compared to the prestudy data. The proportion of annotated features which maintained in the reference target list after curation decreased to 41% (87% in the prestudy). On the one hand, this was caused by the removal of a significant number of redundant features, i.e. keeping features only in the preferable ion mode and discarding the other, and on the other hand, the exclusion of the retention time for scoring during MS-DIAL processing. Using the orthogonal information of the retention time from chromatographic separations beneficially supports the annotation performance. The ratio of confirmed identifications decreased to 33% and the annotation was adjusted only for a proportion of 8%. Here it has to be noted, that 33% represents the number of confirmed identifications in the reference target list, but not completely the annotation performance of MS-DIAL, since correct annotations were removed when belonging to the non-selected ion mode of features being detected in both polarities. This was often the case for glycerophospholipids like (L)PCs and (L)PEs. However, the number of annotated features being reassigned as "Unknown" due to misannotation based on identification scores and retention time pattern was 44% and thus significantly higher in comparison to the prestudy data with around 10%. Most probably, this might be the result of lacking retention time information during the feature identification in MS-DIAL.

As an advantage of targeted data extraction, the target list can be modified as desired by adding lipid targets of interest. Consequently, based on the previous knowledge from the prestudy and former platelet lipidomics investigations [200, 201], additional lipid targets of interest were added to the reference feature list of the main study after manual search for respective peaks and identification based on fragmentation patterns. In detail, the manual identification of lipid features included significantly altered lipids found in the prestudy cohort (e.g. medium chain PCs and oxidized GPs), medium chain GLs, Coenzym species, hydroxy- and oxo-Cholesterol, hydroxy fatty acids, lipids completing a set of lipids from a distinct lipid class (e.g. ACars, TGs, FA), but also additional adducts of internal standards. This set of lipids was compared with the reference list and those lipids were either already present in the reference feature list of the main study, not detected, or added to the target list. Among the latter ones, the identified lipids were either added to the target list as new entries (n=73) or

replaced unknown features which could be identified as one of the lipids ( $n=4$ ). In Table A.4, the additional lipids which were detected, identified, and added to the reference target list are summarized. After the removal of features with a mass accuracy larger than 10 ppm and additional removal of biologically implausible features, i.e. phytoceramides, 900 lipid targets were included in the reference feature list for targeted data extraction.

The identification of lipids in lipidomics research is a challenging and critical step for the generation of biologically confident and relevant study results. Several initiatives and regulatory parties, e.g. Lipidomics Standards Initiative (LSI) [206], Metabolomics Standards Initiative (MSI) [169], the Metabolomics Quality Assurance and Quality Control Consortium (mQACC) [207], or the FDA [146], aim to provide and establish standards to improve the confidence of metabolite identifications. Basically, as recently discussed by Theodoridis et al. [208] and in accordance with the aforementioned initiatives, metabolite identification must involve orthogonal analytical parameters, i.e. accurate mass with fragmentation pattern and retention time or collisional cross section, as done in the present studies (MS and MS/MS data as well as retention time). However, a data curation of all annotations from untargeted LC-MS/MS data is mandatory taking into consideration the analytical and biological plausibility, especially when databases are based on predicted rather than measured spectra [208]. Further, when reporting potential biomarkers, the metabolite identification must be on Level 1, meaning the identity must be verified by an authentic standard [208]. In this clinical platelet lipidomics study, the data curation of lipid annotations was carefully performed after automated data processing and feature identification in MS-DIAL and the number of annotated lipids was significantly reduced, indicating the importance of data curation. Furthermore, after statistical analysis, the identity of significantly altered and highlighted lipids was once again inspected in detail to proof the annotation, potentially improve the identification level by side chain clarification (from lipid species level to molecular species level), and, if possible, standards were used for further confirmation.

#### 4.5.2 Data curation of unknown lipids

As demonstrated, the data curation of identified lipids is a prerequisite for confident lipid reporting. However, also detected features lacking lipid annotation, i.e. unknown features, require curation to avoid the reporting of false-positives. False-positive features might be noise signals and artefacts passing the peak picking criteria during data processing. Furthermore, misalignments of features between samples result in false-positives. To identify and remove false-positive features from the reference feature list, data curation of unknown features was performed for the main study data in terms of a manual curation for misalignments, i.e. inspection of closely eluting peaks being prone for misalignment, and the application of an open-source tool for feature fidelity evaluation named EVA [193].

Firstly, unknown features were inspected for misalignments. Here, misalignment means that two feature entries in the reference list represent the same peak. Features are not cor-

rectly aligned when either the accurate mass or retention time difference between two peaks is larger than the alignment limits. Thus, the ranges for alignment limits must be wide enough to cover the fluctuations and correctly align identical features from different samples. Simultaneously, the ranges must be small enough to avoid the alignment of different features. Consequently, misalignment happens to features with similar accurate masses eluting in a tight retention time range. Therefore, all unknown features were compared in their difference in  $m/z$  ( $\Delta m/z=0.015$ ) and retention time ( $\Delta RT=0.15$ ) and features with both parameters below these limits were inspected in MS-DIAL for similarity or singularity. The results are summarized in Figure 4.18.

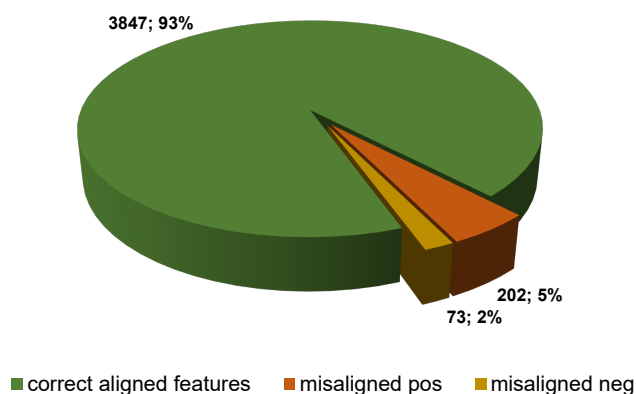


Figure 4.18: Data curation of unknown features: Proportion of misaligned features in positive and negative ion mode is shown individually against the correct alignments combined from both modes.

Among 4122 total unknown features in both positive and negative ion mode, 3847 features (93%) were correctly aligned while 202 (5%) and 73 (2%) were identified as misalignments in positive and negative ion mode, respectively. Misalignments could originate from intra- as well as inter-batch alignment, due to limit exceeding differences in retention time or mass accuracy. It turned out that the latter was the reason for missed alignments in many cases, originating from the initial inter-batch feature alignment of three batches (Batch 06, 07, 15). However, these were singular events for particular features with larger differences in  $m/z$  in comparison of two batches. If a misalignment was detected, the two feature entries were reduced to one. When the reference feature list is applied in targeted data extraction in a later step, the  $m/z$  extraction range ( $\pm 0.01$  Da) is applied on the mean  $m/z$  obtained from the alignment of several batches covering the respective feature in all batches. Exemplary, two unknown features from Batches 06 and 07, being missed to align, are shown in Figure 4.19. The entries in the reference feature list were  $m/z$  873.6243 and  $RT=6.26$  min for Feature 1 and  $m/z$  873.6300 and  $RT=6.30$  min for Feature 2 and the inspection of both peaks in Batch 06 and 07 revealed that the two peaks represent the same feature, thus the alignment was

missed. The Feature 1 was discarded while the second was kept since being more similar with the respective feature in other batches, e.g. Batch 15 with  $m/z$  873.6330. However, in targeted data extraction, the  $m/z$  extraction range of  $\pm 0.01$  Da, i.e.  $m/z$  873.6200 to  $m/z$  873.6400 is sufficient to cover all peaks of the respective feature.

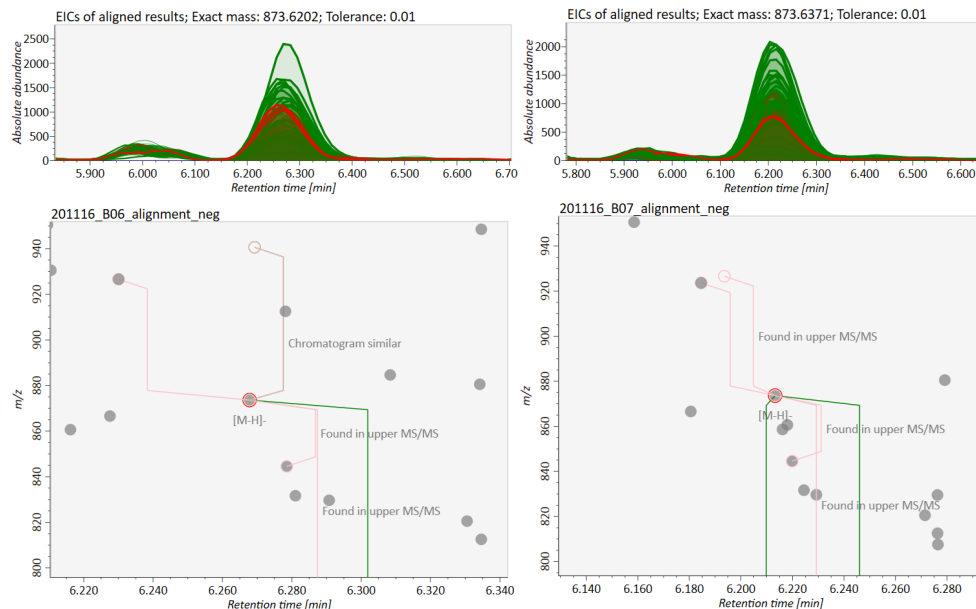


Figure 4.19: Data curation of unknown features: Example of two unknown features ( $m/z$  873.6243, RT=6.26 min and  $m/z$  873.6300, RT=6.30 min in reference list) for which an alignment was missed due to differences in  $m/z$  larger than the applied mass accuracy limit during alignment, i.e.  $\Delta m/z=0.017$  compared to the limit of  $\Delta m/z=0.01$ . The upper panel shows the XICs from the respective peaks in Batch 06 and Batch 07, the panel below the corresponding alignment spotting map both extracted from MS-DIAL.

In addition to misaligned features, unknown features were further curated by the removal of degenerated artefact peaks. For this purpose, the open-source tool EVA was applied. This tool was developed by Guo et al. [193] to evaluate the feature fidelity and is based on a convolutional neural network (CNN) algorithm to identify XICs with poor peak shape being indicative for false-positive features. EVA is open-source and available on the Microsoft Store. To evaluate the feature fidelity of the reference list features, representative samples of all contributing batches from inter-batch feature alignment (eight batches with one sample per batch) were converted into mzXML format in positive and negative ion mode and subjected to the EVA tool. As an output, the user gets the feature list with a "true" or "false" assignment and a XIC plot for all inspected features as well as an over-all prediction summary with the proportion of true and false rated features. In Figure 4.20, an example of a true-positive and false-positive feature is shown as well as the prediction summary over all batches in positive ion mode.

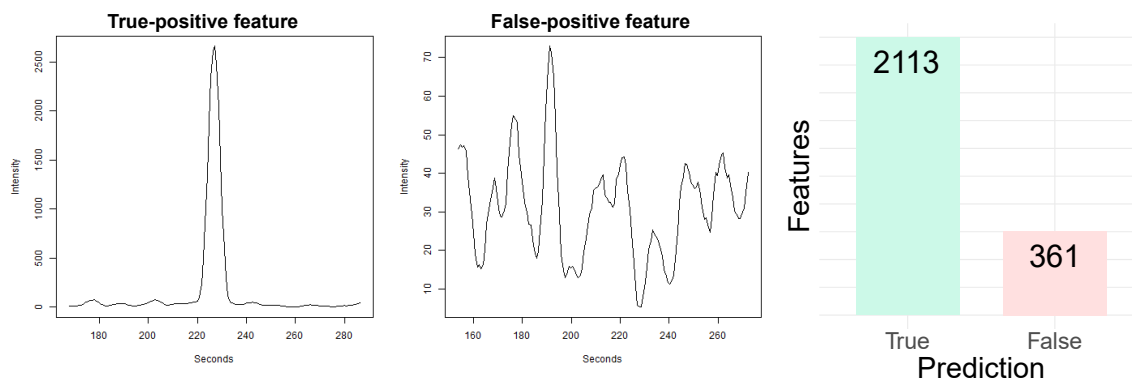


Figure 4.20: Data curation of unknown features: Output of feature fidelity evaluation of unknown features by the EVA tool. Examples of a true-positive and false-positive evaluated feature is shown as well as the prediction summary indicating the proportion of true and false predicted features from Batch 21 positive ion mode.

The evaluation tool had to be applied batch-wise for each ion mode and prediction results were combined from all eight batches to decide whether a feature from the reference list was a true- or false-positive. A feature was discarded as false-positive if it was predicted as false in at least 50% of the batches. As a summary, the proportion of true- and false-positive features among the unknown features from the reference list is depicted in Figure 4.21.

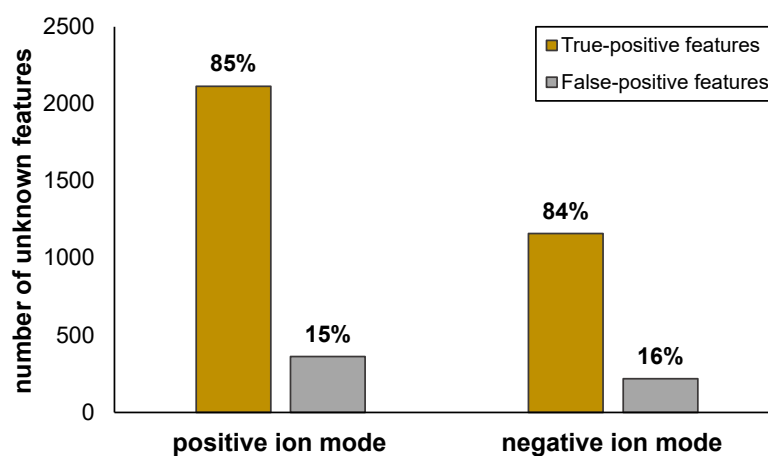


Figure 4.21: Data curation of unknown features: Result of feature fidelity evaluation of unknown features showing the proportions of true- and false-positive features among the unknown features of the reference list in both positive and negative ion mode.

The feature fidelity evaluation revealed 361 features (15%) were false-positive in positive ion mode and 217 (16%) in negative ion mode and were discarded from the reference feature list of the main study for targeted data extraction. These feature fidelity rates are in accordance

with reported rates ranging between 84% and 90% in dependence on data acquisition and data extraction parameters [193]. However, although blank and background subtractions were performed in MS-DIAL, the peak extraction parameters were carefully selected, and a peak count filter was applied to remove random features, a significant proportion of unknown features were degenerated artefact peaks. This result emphasizes the importance of data curation also of unknowns, if those features are intended to be included to statistical analysis due to the detrimental effect of artefact peaks on the statistical outcome.

### 4.5.3 Final reference target list

After curating the features from the reference list obtained from intra- and inter-batch feature alignment, the final target lists were prepared. For the prestudy data, data curation was performed only on annotated lipids since statistical analysis was only based on identified lipids and tools supporting automated fidelity check of unknowns were not available yet, e.g. EVA was published in 2021. To finalize the prestudy reference target list, the lipid notations were adjusted to the lipid notation rules recommended by Liebisch et al. [46]. In total, the target lists were composed of 3156 features with 672 identified lipids in positive ion mode (21% annotation rate) and 953 features in negative ion mode with 282 identified lipids (30% annotation rate) (Figure 4.22).

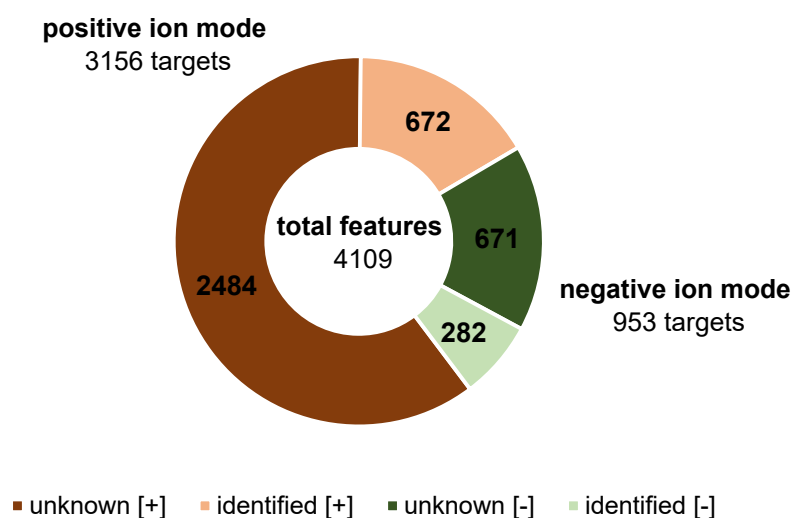


Figure 4.22: Summary of the prestudy reference target list showing the unknown features and identified lipids in positive and negative ion mode.

Among the identified lipids, 16 lipid classes of five lipid categories were covered in positive ion mode, and 20 lipid classes of three categories in negative ion mode based on the LIPID MAPS classification system [44]. The lipid classes and their distribution over the lipid categories are depicted in Figure 4.23.

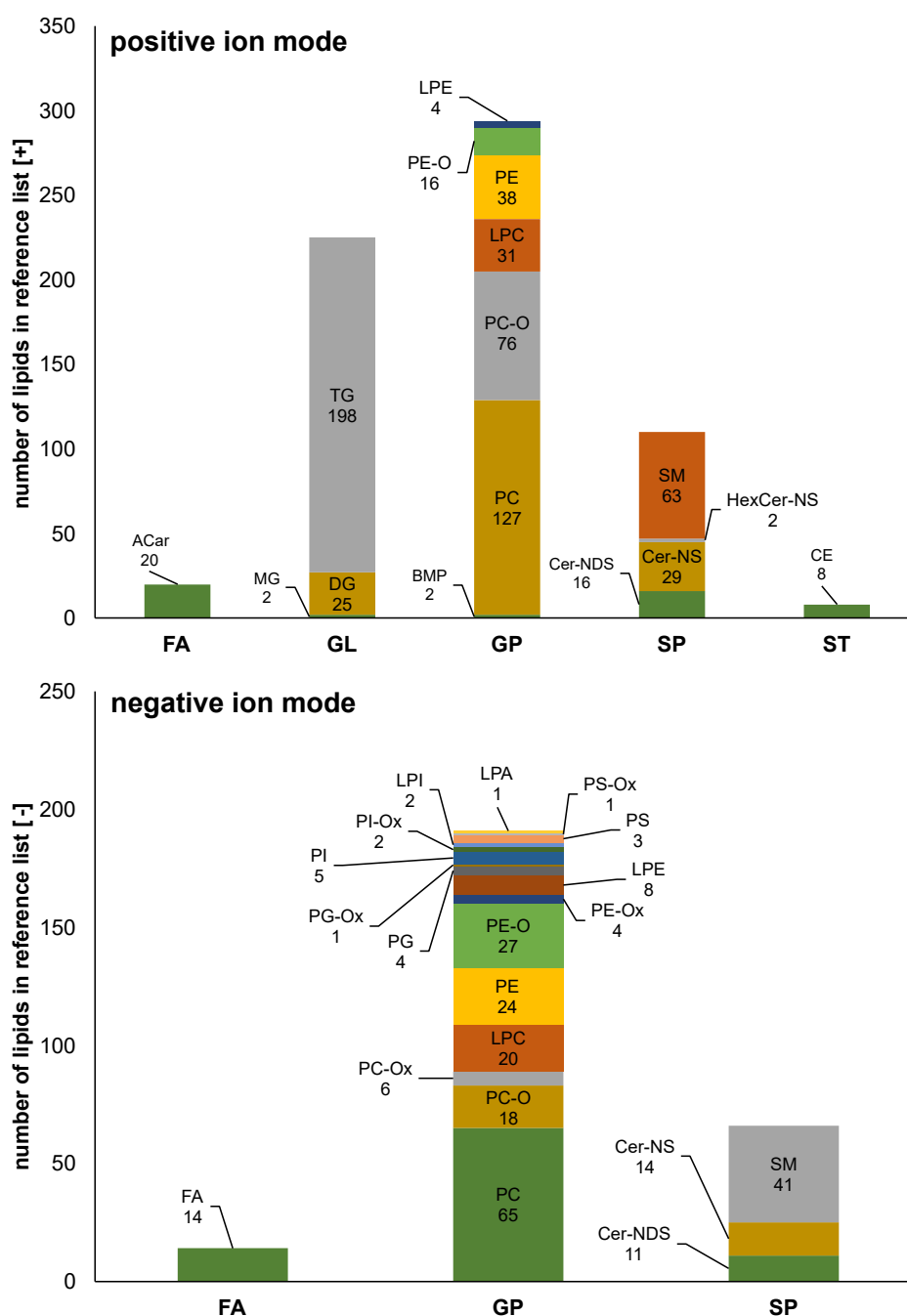


Figure 4.23: Lipid categories and lipid classes of identified lipids of the prestudy reference list. The number of lipids of distinct lipid classes are represented in bars separated into the respective lipid categories based on the LIPID MAPS classification system [44] for positive (top) and negative ion mode (bottom) separately. Internal standards (pos: 15; neg: 11) were not considered in this representation. FA=Fatty Acyls, GL=Glycerolipids, GP=Glycerophospholipids, SP=Sphingolipids, ST=Sterol Lipids.

The final reference target lists for the prestudy cohort, one for each polarity mode, contained the alignmentID, the metabolite name, the precursor ion mass, the reference retention time, i.e. mean RT after alignment, and the peak extraction range, i.e.  $\pm 0.01$  Da of the precursor ion mass.

For the finalization of the main study reference target list, the reference feature list of identified lipids from the prestudy cohort was aligned against the entire reference list from the main study data to harmonize the identified lipids from both cohorts. After retention time adjustment, 663 features in the positive ion mode and 276 features in negative mode (all identified lipids after exclusion of internal standards) were aligned against 3388 and 1518 features in the respective polarity mode of the main study. A result summary is given in Figure 4.24.

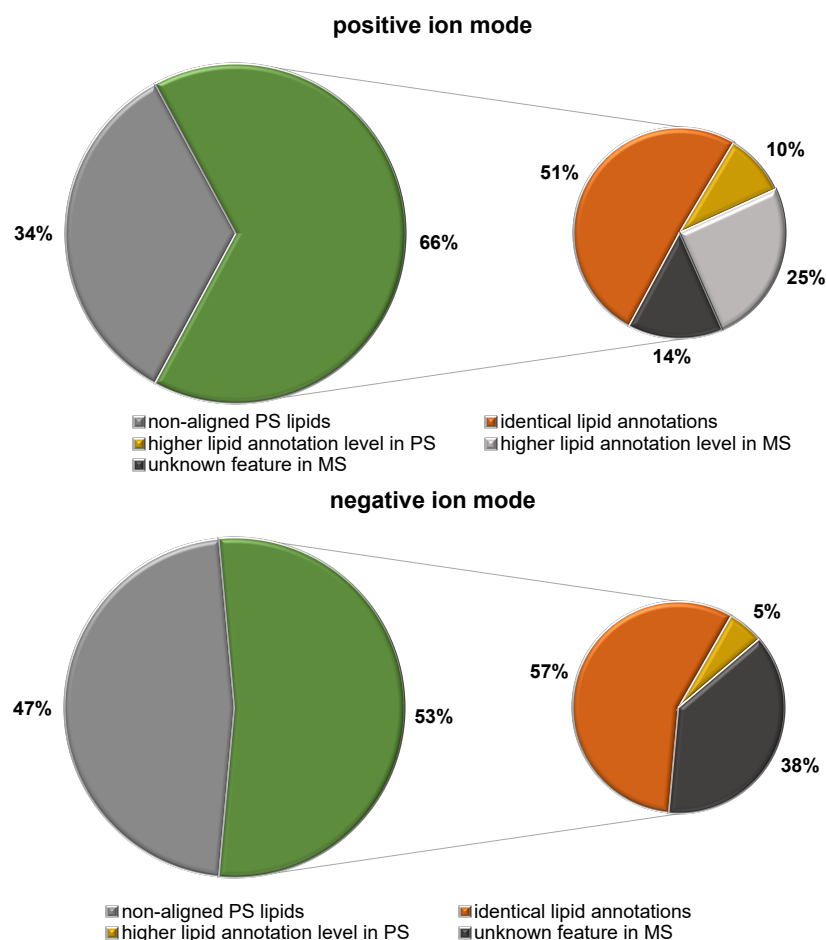


Figure 4.24: Result summary of feature list alignment from prestudy (PS) and main study (MS) datasets. For both positive (top) and negative (bottom) ion mode, the large pie shows the proportion of aligned and non-aligned lipids between pre- and main study among the features of the prestudy list. The small pie further differentiates the aligned features based on the annotation: identical annotations in PS and MS, higher annotation levels in PS or MS, and unknown features in MS identified in PS.



Here, two-third (66%) and half (53%) of the identified lipids of the prestudy feature list were aligned with the main study features in positive and negative ion mode, respectively. Among them, annotations were identical in both pre- and main study (51% and 57% in positive and negative ion mode), while also a significant proportion of unknown lipids could be annotated through the alignment (pos: 63 features (14%); neg: 55 features (38%)). For the remaining alignments, differences in identification levels were observed with either a higher level in the prestudy (pos: 10%; neg: 5%) or in the main study (pos: 25%) cohort. The proportion of non-aligned lipids from the prestudy reference list seems to be relatively high (pos: 34%; neg: 47%). This is caused by the removal of lipids in one polarity mode if ionizing in both in the main study. For example, PCs were detected in both polarity modes. They were preferentially kept in the positive ion mode reference list while being discarded from the negative ion mode list. Since the number of PC species was high in both polarity modes, this contributed to the high ratio of non-aligned lipids in the negative ion mode. Finally, the number of annotated lipids in the main study reference list was increased by 67 features, i.e. 56 in positive and 11 in negative ion mode, and unknown features were reduced by 114. Due to ionization redundancies and the removal of respective features, the decrease in unknown features was larger than the increase in annotated lipids.

Through the entire multi-step process of data curation, the number of features in the reference feature list of the main study was significantly reduced. The decrease throughout several cleaning steps is shown in Figure 4.25.

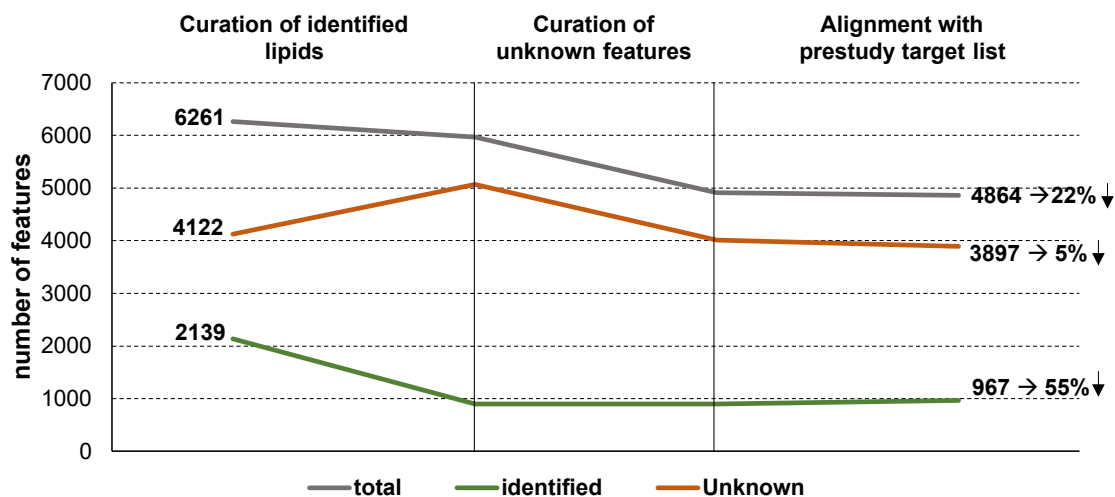


Figure 4.25: Decrease of features in reference list through data curation. The number of features is plotted after the respective curation step starting after inter-batch feature alignment followed by curation of identified lipids including manual addition of lipid targets, curation of unknowns, alignment with prestudy reference list up to the final number of features in the reference target list for the main study cohort. The progression is shown for all, unknown, and identified features.

The data curation resulted in an over all decrease of features of finally 22%. The number of unknown features increased after the curation of annotated features since misannotated but true-positive features were reassigned as unknown. Further unknown feature curation caused a reduction of those features of finally 5% compared to prior curation. The strongest decrease was observed for the annotated lipids (55%) since those were reduced by misannotated lipids, false-positive features, and due to the removal of ionization redundancies. The final reference target list was composed of 4864 features among which 967 were annotated lipids with an annotation rate of 20%. In Figure 4.26, the detected lipid classes with the respective proportions of lipid species are plotted. The 967 identified lipids were distributed over 6 lipid categories and 40 lipid classes.

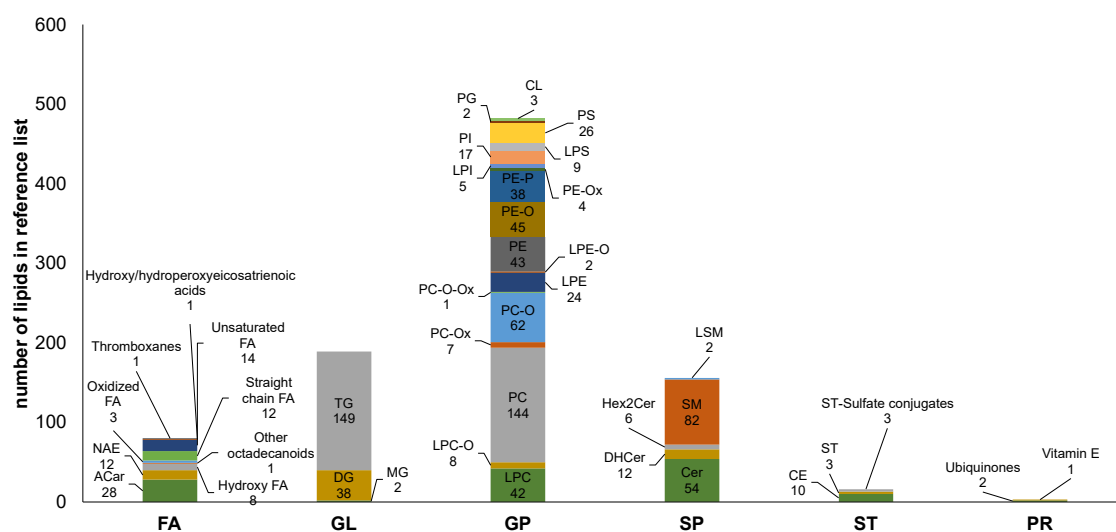


Figure 4.26: Lipid categories and lipid classes of identified lipids of the main study reference list. The number of lipids of distinct lipid classes are represented in bars separated into the respective lipid categories based on the LIPID MAPS classification system [44]. Internal standards (total: 41) were not considered in this representation. FA=Fatty Acyls, GL=Glycerolipids, GP=Glycerophospholipids, SP=Sphingolipids, ST=Sterol Lipids, PR=Prenol Lipids.

Finally, the combined reference list was distributed into two target lists, one for each polarity mode, for targeted data extraction of the main study cohort. The reference target lists for the positive ion mode with 3380 features and for the negative ion mode with 1476 features included the following information: alignmentID, polarity mode, metabolite name on lipid species and molecular lipid species level, if available, the precursor adduct type, the precursor exact mass, the reference retention time, the metabolite name for data extraction in MultiQuant (concatenated information of metabolite name, adduct type, and alignmentID), and the peak extraction range, i.e. precursor exact  $m/z \pm 0.01$  Da. For specific lipid features, several peak extraction ranges were listed. In detail, sum extraction was performed for internal standards with consideration of multiple precursor adducts and isotopologues ( $[M-2]$ ,  $[M-1]$ ,  $[M+1]$ ,

[M+2]), as well as for Coenzyme Q10(ox) considering Coenzyme Q10(red) due to in-source reduction.

## 4.6 Preparation of the large-scale lipidomics datamatrix

After finalization of the reference target list, the further steps of data processing were dedicated to prepare the lipidomics datamatrices for both study cohorts. Firstly, based on the cohort-specific reference lists, targeted data extraction was performed for all batches of pre- and main study datasets and respective datamatrices were prepared. Further, for the selection of the optimum data normalization strategy, several normalization methods were applied and evaluated for their performance prior applying the best method to the pre- and main study data. The final lipidomics datamatrices were obtained after additional cleaning steps and the merging of lipidomics with clinical data. These procedures are explained in detail in the following three sections.

### 4.6.1 Targeted data extraction

The initial step of the targeted data extraction procedure was the adjustment of the reference retention times specified in the reference target list to the batch-specific retention times. Similar to the retention time adjustment prior inter-batch feature alignment, the retention times of a subset of features were determined for each batch by data extraction with MultiQuant and plotted against the reference value. Based on the linear correlation, the retention times of all features in the reference target list were adjusted for each batch. In the main study processing, the subset of features was modified compared to the set used prior the alignment to improve the elution range coverage, i.e. from  $\sim 2$  features/min to  $\sim 2.5$  features/min. An exemplary representation of the correlation of reference and batch-specific retention time is given in Figure 4.27.

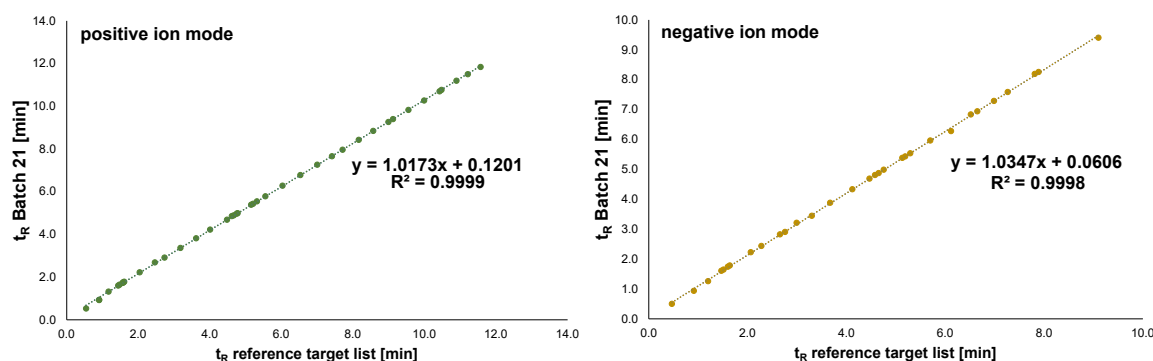


Figure 4.27: Retention time adjustment from reference to batch-specific values shown for Batch 21 in positive (left) and negative (right) ion mode. The correlation was based on 41 features in positive and 31 features in negative ion mode covering the entire relevant elution range.

The linear equations for retention time adjustment of all batches are summarized in Table A.5. To assess the performance of the retention time prediction in each batch, the accuracy, i.e. rating the mean detected to the expected retention time, as well as the precision were determined for each feature of the reference target list in positive and negative ion mode. In Figure 4.28, the retention time accuracy and precision is depicted for the eight batches of the main study, which were used in inter-batch feature alignment. As illustrated by the bar charts, the accuracy was ranging between 94% and 108% with a gaussian curve shape exhibiting its maximum at 100% and the precision was always below 10% with the maximum of features having a CV below 1%, both confirming the proper performance of the retention time adjustment.

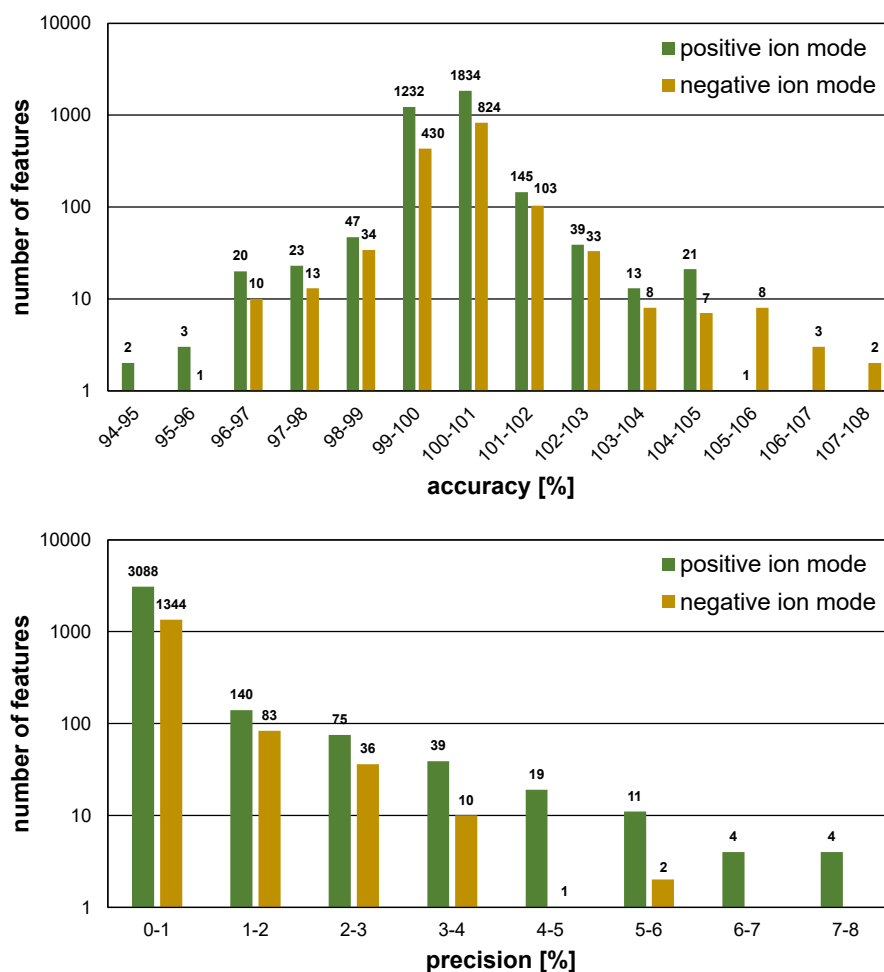


Figure 4.28: Performance of the retention time adjustment shown as accuracy and precision of all features in positive (green) and negative (yellow) ion mode for the eight main study batches being used for inter-batch feature alignment. Accuracy was determined as the ratio of mean detected vs. expected retention time represented as the mean accuracy over eight batches. The precision was calculated as the ratio of the standard deviation of the detected retention time within a batch against the batch-specific detected mean RT and is represented as the mean precision over eight batches.

Subsequently, the reference target lists prepared for the pre- and main study in positive and negative ion mode with batch-specific retention times were used for targeted data extraction in Sciex MultiQuant. For data integration, the MQ4 algorithm was applied and the method parameters are summarized in Table A.6. The MultiQuant results were exported in *txt*-format including the following qualitative and quantitative parameter: sample name, component name, expected RT, area, height, quality, retention time, width at 50%, S/N, tailing factor, asymmetry factor. The results from all batches were combined in one datamatrix per polarity mode.

These raw datamatrices required further curation in terms of feature detection rate assess-

ment as well as missing value imputation. Missing values are often observed in untargeted metabolomics datasets, especially in large-scale studies, arising due to biological, technical, or computational reasons [153–155, 196]. Missing values have detrimental effects on the statistical analysis of untargeted lipidomics, thus, the imputation of those values is essential for further analysis.

Firstly, the feature detection rates were determined to identify rarely detected features among all batches and samples, with high proportions of missing values indicating analytical issues. Hence, the number of missing values was evaluated for all features in all clinical and QC samples among the entire datasets of pre- and main study. The overall proportion of missing values (Figure 4.29) was relatively low in the prestudy dataset (2% in each positive and negative ion mode) and in the main study dataset (11% and 18% in positive and negative ion mode, respectively) compared to 10-30% as reported in literature [154, 155, 196]. This might be caused by the targeted data extraction manner allowing lower intensity thresholds for quantification than typically applied in untargeted data extraction procedures.

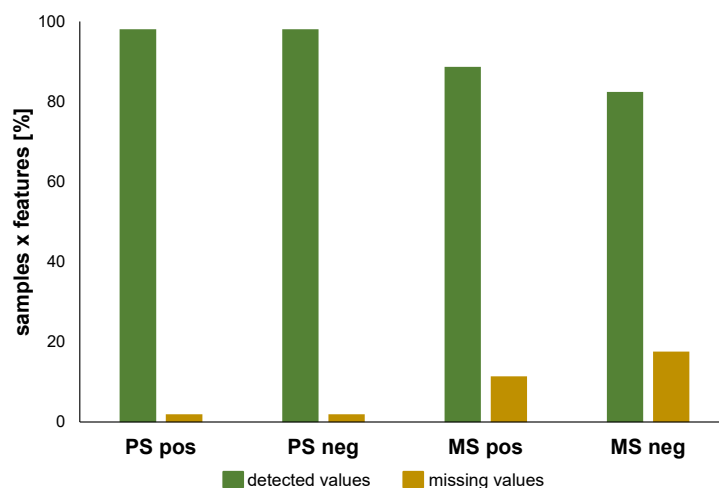


Figure 4.29: Proportion of missing values among all samples and features in prestudy (PS) and main study (MS) data in positive (pos) and negative (neg) ion mode, respectively.

Based on the missing value ratios, features and samples with low detection rates were discarded from further processing. Features were removed when not being detected in all batches, i.e. three in the prestudy and 22 in the main study, and additionally when the detection rate over all samples was below 50%. These limits were defined to avoid detrimental effects of rare feature detection or absence of a feature in an entire batch on the further processing, i.e. missing values imputation and data normalization, as well as statistical analysis. Thus, in the prestudy, 176 features were discarded in the positive ion mode while 3 features were removed in the negative ion mode. For the main study dataset, it was 187 and 342 in positive and negative ion mode, respectively. Further, in the prestudy dataset, 13 samples had to be removed due to high proportions of missing values, i.e. >80%, in at least one polarity mode, and 14 samples (10 clinical and 4 QC samples) were excluded in the main study dataset. Missing values were then imputed by the half-minimum method for the prestudy data, thus, missing values were replaced by a value being half of the detected minimum response over all samples. Caused by data processing optimization, random forest (RF) algorithm was applied in the main study datamatrices using RStudio [153, 194]. The RF algorithm is highly suitable for missing values imputation in untargeted profiling data and is a non-parametric method, thus no group assignment in advance is required [153].

Subsequently to missing values imputation, multiple precursor ion adducts of a lipid species were combined and type I isotopic correction was performed for the annotated lipids in the main study data processing. The summation had the goal to quantify a lipid species instead of a single precursor ion adduct and was done for individual lipid species of Cers and Sterols ( $[M+H]^+$ ;  $[M+H-H_2O]^+$ ), DGs and TGs ( $[M+NH_4]^+$ ;  $[M+Na]^+$ ), (L)PCs and SMs ( $[M+H]^+$ ;  $[M+Na]^+$ ). The adduct addition was only performed in positive ion mode due to the absence of multiple precursor ion adducts in negative ion mode. For 118 lipid species, two precursor ions were summed together.

Further, to quantify for the total ion abundance, type I isotopic correction was performed. In dependence on the carbon number of a lipid species, isotopic effects from  $^{13}C$  abundance contribute to differences in the ion intensity of  $[M]$  [156]. The application of corrective calculations contributes to more accurate lipid quantification since quantitative results with and without isotopic correction differ significantly in relative quantifications where the single lipid class specific standard has a significantly different carbon number, i.e. up to around 50%, especially for lipids with high carbon numbers [156]. Thus, for consideration of all isotopologues ( $[M]$ ,  $[M+1]$ ,  $[M+2]$ ) in quantification, type I isotopic correction was applied by calculation to all annotated lipids (Formula 3.2 in Section 3.2.2.5).

For internal standards, both multiple adduct summation and isotopic correction was not required since those features were already sum extracted during targeted data extraction in MultiQuant, i.e. all detected precursor ion adducts and five isotopologues were quantified together.

### 4.6.2 Data normalization

After curating the datamatrices containing raw data, the optimal data normalization strategy was selected through testing and evaluating several normalization methods. Data normalization is mandatory especially in large-scale omics studies to correct the raw data for signal drifts to improve the statistical analysis, thus the study output. For this purpose, different normalization methods were applied on subsets of raw data, i.e. positive ion mode of all three prestudy batches and eight batches from the main study (batches from inter-batch feature alignment), respectively. It is advisable to optimize the data normalization strategy for each dataset individually, since normalization performance depends on the acquired data [180]. Due to differences in sampling and sample preparation, albeit of minor extent, normalization evaluation was performed for both cohorts separately. Different sets of methods were used in pre- and main study datasets among which normalization was based on internal standards (CCMN, NOMIS, B-MIS, RUVrandom) [158, 177, 181, 182], QC samples (LO(W)ESS, SERRF, SVM) [159, 184–187], sample-specific values (PCnorm, PQN, mTIC) [174–176], or on a combination of two different methods. While nearly all methods were applied in RStudio by using published R scripts from Drotleff et al. [180] and Fan et al. [159], the sample-specific scaling method PCnorm was based on an idea of Chao et al. [114] who used a precursor ion scan of the specific phosphocholine head group fragment ( $m/z$  184) for normalization in lipidomics. In the present study, all lipid species belonging to the PC lipid class were summed up to a sample-specific scaling factor and every feature in a sample was normalized by this value. Since PCs are the major membrane building lipid class, they are supposed to provide a stable and representative sample-specific scaling factor in accordance with the findings from Chao et al. [114]. However, also SM produce the specific fragment of  $m/z$  184, but they were not considered for the scaling factor calculations due to their regulatory role in inflammation.

For the evaluation of the best normalization strategy in the prestudy data, several normalization methods were applied to the three batches in positive ion mode including 107 clinical and 33 QC samples. Clinical samples were separated into groups based on the diagnosis of acute coronary syndrome (ACS,  $n=26$ ) and chronic coronary syndrome (CCS,  $n=81$ ). QC-based, IS-based normalization methods as well as a combination of SERRRF and RUVrandom methods were applied and the variation metrics, i.e. coefficient of variation (CV), the median absolute deviation (MAD), and the variance (Var), were determined as published by Drotleff et al. [180] (Figure 4.30).



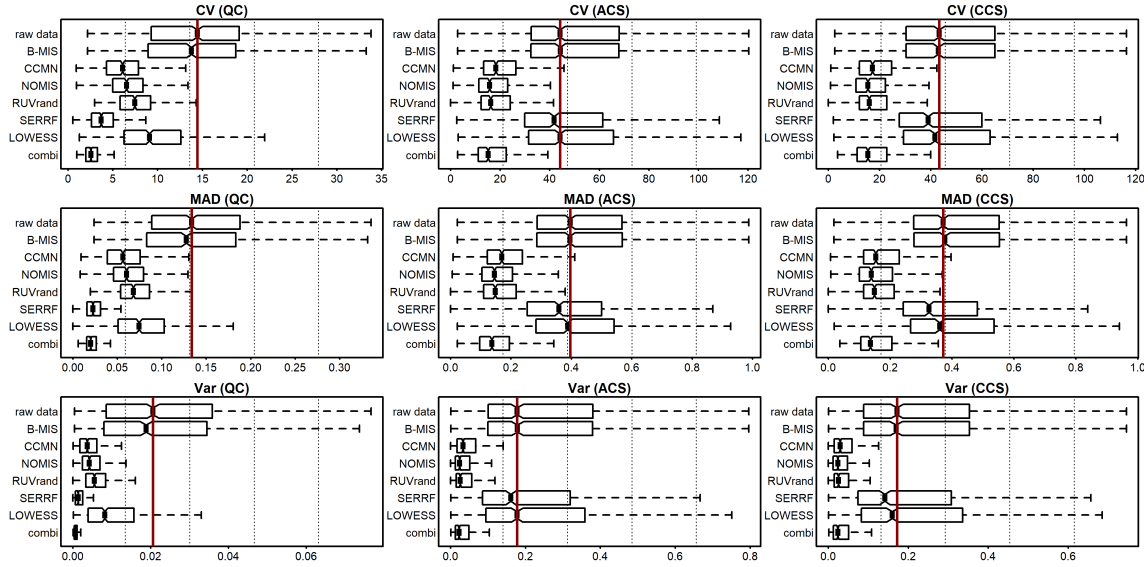


Figure 4.30: Overview of metrics of variation for raw and normalized data in box-whisker plots in prestudy data. The red line marks the median value of raw data in coefficient of variation (CV; top line), median absolute deviation (MAD; medium line), and variance (Var; bottom line). Combi: SERRF-RUVrandom.

Through the application of both QC- and IS-based normalization methods, the variation of feature responses was reduced within the sample groups (QC, ACS, CCS). The variation decreased in terms of the median values of CV, MAD, and Var compared to the raw data (position left from the red line) as well as in the spreading around the median (size of the boxes). While B-MIS did not result in a significant decrease of variation, data were successfully normalized by the other IS-based methods, also being superior to QC-based methods in terms of variation reduction in the clinical sample groups. The best normalization performance was obtained by a combination of QC- and IS-based methods using SERRF and RUVrandom algorithms, thus, this strategy was selected for the application to the entire dataset.

Finally, the inter-batch coefficient of variation (CV) was determined for raw and normalized data (Figure 4.31) in QC samples. Through data normalization by a combination of QC-based SERRF method and IS-based RUVrandom method, the inter-batch variation was significantly reduced and nearly all features (98%) exhibited a  $CV < 30\%$ , i.e. FDA recommendation for precision in biomarker research.

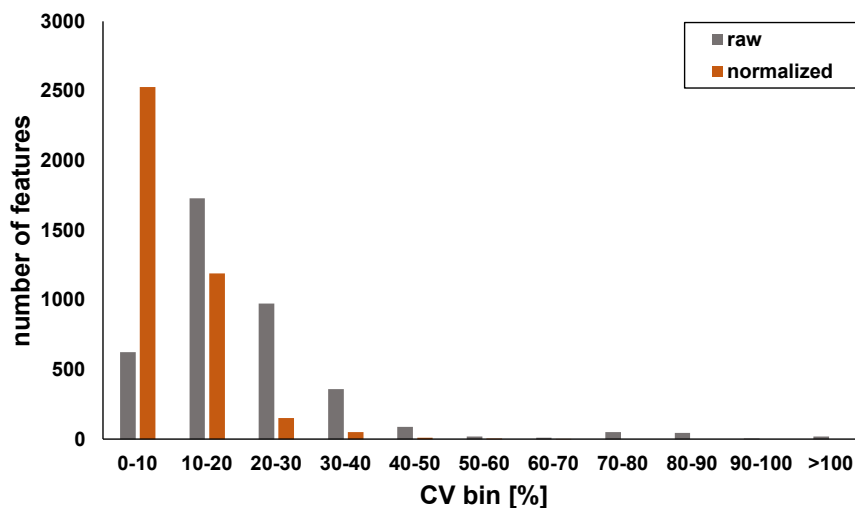


Figure 4.31: Reduction of inter-batch CV by data normalization compared to raw data determined in QC samples of prestudy data. Data normalization by a combination of SERRF and RUVrandom methods was applied on the entire dataset in positive and negative ion mode.

The normalization method selection in the main study dataset was based on the positive ion mode data from eight batches including 387 clinical and QC samples with sample group assignment to ACS ( $n=128$ ), CCS ( $n=171$ ), and QC ( $n=88$ ). However, 110 samples from 497 in the eight batches were excluded from the normalization optimization due to unclear clinical group assignment or as NIST samples. Similar to the procedure in the prestudy, the CV, MAD, and Var were determined after application of different normalization strategies and are presented in Figure 4.32.

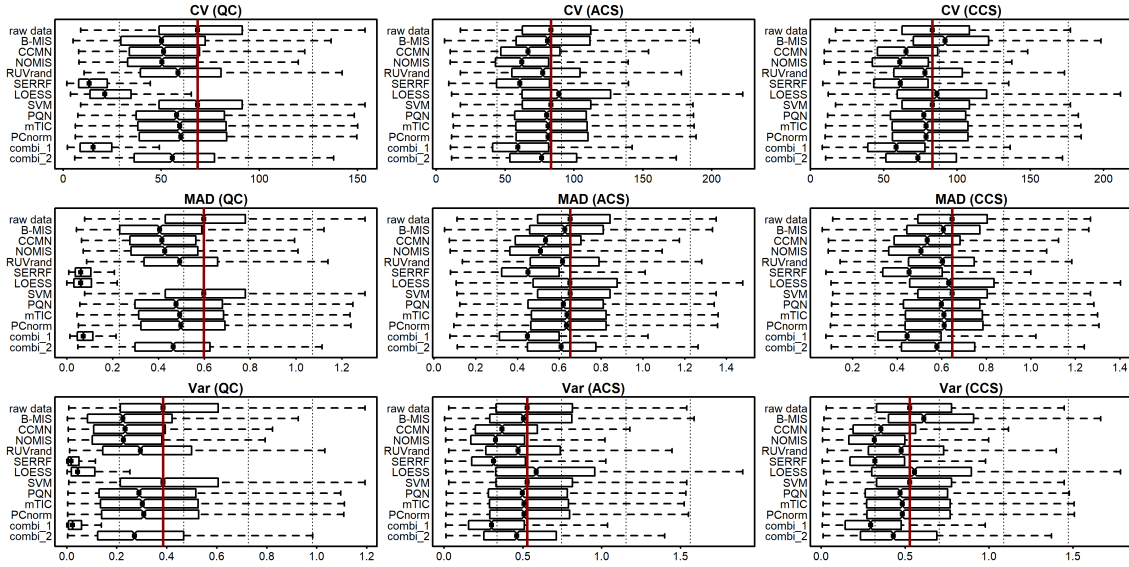


Figure 4.32: Overview of metrics of variation for raw and normalized data in box-whisker plots in main study data. The red line marks the median value of raw data in coefficient of variation (CV; top line), median absolute deviation (MAD; medium line), and variance (Var; bottom line). Combi\_1: SERRF-CCMN, Combi\_2: mTIC-CCMN.

The assessment of variation metrics revealed that nearly all normalization methods in all investigated groups led to a decrease in intra-group variation, except of LOESS in ACS and CCS groups. The best performance was obtained by the QC-based SERRF normalization in terms of the largest reduction of all CV, MAD, and Var in QC as well as in clinical sample groups. The performance of the method combination of SERRF and CCMN was similar to SERRF alone. Furthermore, the evaluation showed that IS-based methods outperformed the sample-based strategies especially in clinical sample groups. Here, in addition to variation metrics, multivariate statistics was used for further normalization performance assessment, i.e. partial least-squares discriminant analysis (PLS-DA). The model parameters cumulative goodness of fit ( $R^2Y$ ), cumulative goodness of prediction ( $Q^2Y$ ), and number of optimum components for PLS-DA is given in Table 4.10.

Table 4.10: Summary of PLS-DA results for raw and normalized data

| normalization method |            | MVDA   |        |            |
|----------------------|------------|--------|--------|------------|
|                      |            | $R^2Y$ | $Q^2Y$ | #Opt.Comp. |
| -                    | raw data   | 0.976  | 0.745  | 10         |
| IS-based methods     | B-MIS      | -      | -      | -          |
|                      | CCMN       | 0.943  | 0.611  | 8          |
|                      | NOMIS      | -      | -      | -          |
|                      | RUVrandom  | 0.976  | 0.768  | 10         |
| QC-based methods     | SERRF      | 0.975  | 0.572  | 6          |
|                      | LOESS      | 0.970  | 0.703  | 9          |
|                      | SVM        | 0.976  | 0.745  | 10         |
| sample-based methods | PQN        | 0.973  | 0.729  | 10         |
|                      | mTIC       | 0.148  | 0.079  | 1          |
|                      | PCnorm     | 0.149  | 0.079  | 1          |
| 2-step normalization | SERRF-CCMN | -      | -      | -          |
|                      | mTIC-CCMN  | 0.943  | 0.718  | 8          |

With the lowest number of required components (6 components), SERRF-normalized data provided a good performance in terms of goodness of fit ( $R^2Y=0.975$ ) and goodness of prediction ( $Q^2Y=0.572$ ) compared to raw data ( $R^2Y=0.976$ ;  $Q^2Y=0.745$ ; components=10). In the group of IS-based methods, CCMN performed best in terms of model parameters and number of components ( $R^2Y=0.943$ ;  $Q^2Y=0.611$ ; components=8), among the sample-based methods it was PQN ( $R^2Y=0.973$ ;  $Q^2Y=0.729$ ; components=10). The PLS-DA of SERRF-normalized data is shown in Figure 4.33 against raw, CCMN-, and PQN-normalized data.

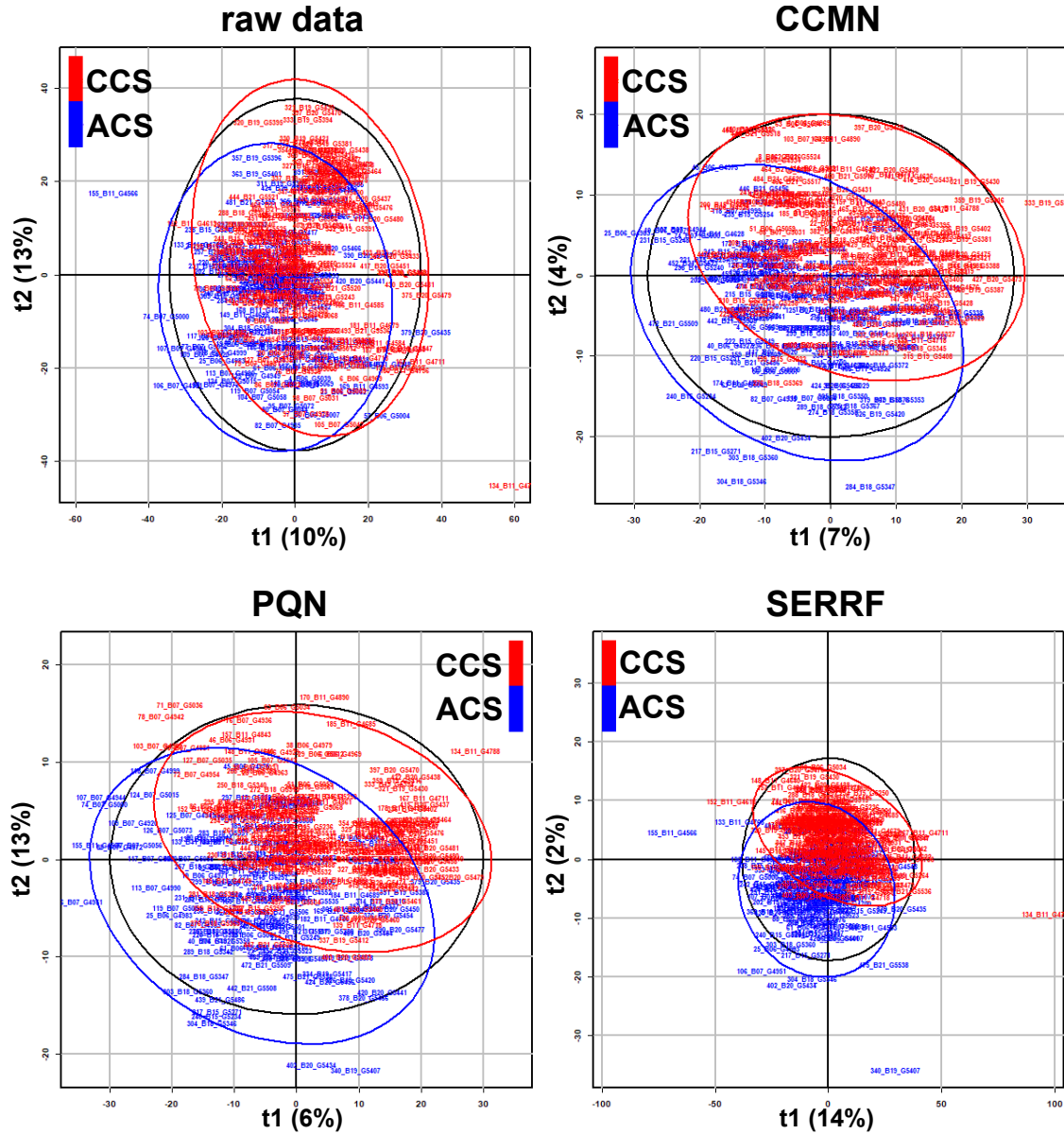


Figure 4.33: PLS-DA plots of raw and CCMN-, PQN-, and SERRF-normalized data. The model parameter are summarized in Table 4.10.

In all PLS-DA plots, a slight separation of samples due to disease severity (ACS vs. CCS) was observed. All shown normalization methods contributed to a better inter-group separation, among which only SERRF normalization led to a significant reduction of intra-group variation. Thus, under consideration of variation metrics and the performance in multivariate statistics, the SERRF method was selected as the optimum normalization strategy for the entire main study dataset.

Thus, SERRF normalization was applied to the entire dataset in positive and negative ion mode. In addition, CCMN normalization was performed for comparison. For raw data as well

as SERRF- and CCMN-normalized data, the inter-batch CV of all features in QC samples was determined and the results are shown in Figure 4.34. The CV assessment confirms the conclusion, that only by SERRF normalization the inter-batch variation of all features could be reduced significantly reaching the limit of  $CV < 30\%$  for the majority of features (96%). Due to a  $CV > 30\%$  in QC samples after normalization, 161 features were finally discarded in raw and normalized datasets.

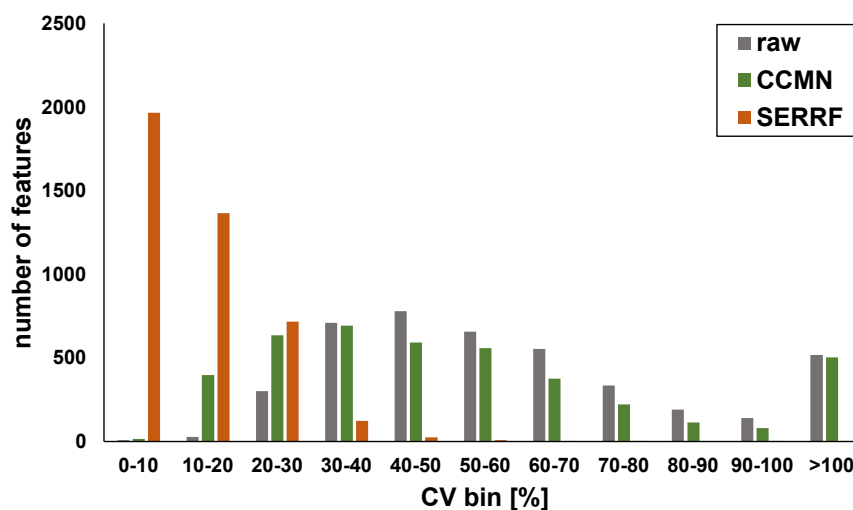


Figure 4.34: Reduction of inter-batch CV by data normalization compared to raw data determined in QC samples of main study data. Data normalization by CCMN and SERRF methods was applied on the entire dataset in positive and negative ion mode. In the CCMN-normalized data, internal standards were excluded from CV evaluation.

As indicated by the individual evaluation of data normalization for pre- and main study datasets, the method performance of the same methods differed between the cohorts. As becoming evident from the literature, the proper method selection for data normalization depends on the acquired dataset. For example, while B-MIS normalization performed best in the plasma lipidomics study presented by Drotleff et al. [180], QC-based methods, especially SERRF, outperformed B-MIS normalization considerably in the lipidomics investigations by Fan et al. [159]. In untargeted studies, QC-based strategies have the advantage that each feature is normalized with the responses of the respective feature in QC samples, thus, under identical analytical conditions, e.g. retention time and matrix composition in the ion source. However, the sample matrix of QC samples must be similar with the matrix of the clinical samples and identical in all acquisition batches. Therefore, the QC samples for the entire study have to be prepared in advance, e.g. leading to issues in QC sample matrix availability or storage. In contrast, IS-based methods have the advantage that the normalization agent is spiked to each sample during sample processing without the necessity of additional sample preparation or sample matrix limitations. Moreover, differences in extraction recoveries might be corrected since the IS are usually spiked prior lipid extraction. However, the number of

lipid features highly exceeds the manageable amount of internal standards, i.e. thousands of features normalized by a few tens of IS, thus, lacking coelution and similar matrix effect for IS and lipid features. The limitations of QC- and IS-based strategies are overcome by sample-based methods, in which features are normalized by a sample-specific value, but here without an advantage in normalization performance. Hence, based on the normalization evaluation of the pre- and main study lipidomics data, it is advantageous to include all opportunities of data normalization in the study design and evaluate the optimum method after data acquisition.

### 4.6.3 Final lipidomics datamatrix

To finalize the complex data processing procedure (Figure 4.4), the data from the two polarity modes were combined and raw and normalized data each were hyphenated with the clinical data. Moreover, lipid concentrations were determined by relative quantification with internal standards. After discarding features with a  $CV > 30\%$  in normalized QC samples (see Section 4.6.2), the number of finally included features in the datamatrices of prestudy and main study cohorts is given in Table 4.11.

Table 4.11: Overview of all features in the final datamatrices of prestudy and main study cohorts

| features          | prestudy |           | main study |           |
|-------------------|----------|-----------|------------|-----------|
|                   | absolute | ratio [%] | absolute   | ratio [%] |
| <b>identified</b> | 939      | 24        | 767        | 19        |
| <b>all</b>        | 3915     | 100       | 4013       | 100       |

Both lipidomics matrices contained around 4000 features with a lipid identification rate of 24% in the prestudy and 19% in the main study. The number of identified lipids was larger in the prestudy cohort since a removal of redundant entries of lipids ionizing in both polarity modes only became part of the processing procedure during workflow optimization based on the main study data.

In addition to the raw and normalized data, lipid concentrations were calculated for identified lipids based on a relative quantification approach using the corresponding lipid class specific internal standards. Relative quantification by isotopically labeled or odd-chain internal standards is usually applied in untargeted lipidomics due to the absence of lipid species matching standards [58]. Through one-point calibration ( $\text{concentration}_{\text{lipid}}[\text{pmol}/10^6 \text{ PLTs}] = \text{intensity}_{\text{lipid}}/\text{intensity}_{\text{IS}} * \text{concentration}_{\text{IS}}[\text{pmol}/10^6 \text{ PLTs}]$ ) with the respective lipid class specific IS, identified lipids were quantified in molar concentrations related to the cell count ( $\text{pmol}/10^6 \text{ PLTs}$ ). In case of a missing lipid class specific IS, a surrogate IS was selected originating from the same lipid category and/or exhibiting similar retention time (Table 4.12).

Table 4.12: Assignment of IS to lipid classes for quantification by one-point calibration with lipid class specific IS

| lipid category | IS                                                                                                                                                         | main study<br>quantification of                                                                               | IS                                                                                                                                     | prestudy<br>quantification of                                                                         |
|----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| <b>FA</b>      | CAR 16:0(d3)<br>FA 20:4(d11)                                                                                                                               | ACar, NAE<br>FA classes                                                                                       | FA 20:4(d8)                                                                                                                            | FA classes                                                                                            |
| <b>GP</b>      | LPC 18:1(d7) <sup>1</sup><br>PC 15:0_18:1(d7)<br>LPE 18:1(d7) <sup>1</sup><br>PE 15:0_18:1(d7)<br>PI 15:0_18:1(d7)<br>PG 15:0_18:1(d7)<br>PS 15:0_18:1(d7) | LPC, LPC-O<br>PC, PC-O, PC-Ox, PC-O-Ox, CL<br>LPE, LPE-O, LPI, LPS<br>PE, PE-O, PE-P, PE-Ox<br>PI<br>PG<br>PS | LPC 18:1(d7) <sup>1</sup><br>PC 15:0_18:1(d7)<br>LPE 18:1(d7) <sup>1</sup><br>PE 15:0_18:1(d7)<br>PI 15:0_18:1(d7)<br>PG 15:0_18:1(d7) | LPC, LPA, ACar, BMP<br>PC, PC-O, PC-Ox, PS-Ox<br>LPE, LPI<br>PE, PE-O, PE-P, PE-Ox<br>PI<br>PG, PG-Ox |
| <b>GL</b>      | MG 18:1(d7)<br>DG 15:0_18:1(d7)<br>TG 15:0_18:1(d7)_15:0                                                                                                   | MG<br>DG<br>TG                                                                                                | DG 15:0_18:1(d7)<br>TG 15:0_18:1(d7)_15:0                                                                                              | MG, DG<br>TG                                                                                          |
| <b>SP</b>      | SM 18:1;2O/18:1(d9)<br>LSM 18:1;O2(d9)                                                                                                                     | SM<br>LSM                                                                                                     | SM 18:1;2O/18:1(d9)<br>Cer 18:1;O2/18:0(d7)                                                                                            | SM<br>Cer-NS, Cer-NDS, HexCer-NS                                                                      |
| <b>ST</b>      | Cer 18:1;O2/15:0(d7)<br>ST 27:1;O(d7)<br>CE 18:1(d7)                                                                                                       | Cer, DHCer, Hex2Cer<br>ST<br>CE                                                                               | CE 18:1(d7)                                                                                                                            | CE                                                                                                    |

<sup>1</sup> Quantification using both detected isomers.

To conclude, for pre- and main study cohort, three datamatrices each were obtained being composed of clinical data hyphenated with i. raw data of all features, ii. normalized data of all features, and iii. lipid concentrations of identified lipids. These lipidomics data were further subjected to statistical analysis to elucidate the role of the platelets lipid profile in different clinical and biological investigations.

## 4.7 Considerations for further methodological and workflow optimization

The successful completion of a large-scale lipidomics study might be the start of further large-scale investigations, once the challenges of large sample sizes compared to limited sample numbers are overcome. Thus, based on the experiences from the data acquisition and processing of the two presented studies, there are some opportunities for methodological and workflow optimization.

Regarding the methodological part including sample preparation, batch design, and LC-MS data acquisition, optimization in IS spiking, QC samples, and LC-MS method might be advantageous. An extension of the set of spiked internal standards might be reasonable to account for the structural diversity of lipids and improve the chemical representation of lipids by IS. The application of several IS per lipid class differing in carbon chain length, degree of saturation, and bond types can contribute to an improvement in lipid quantification since the chemical structure of lipids effect the MS response [156].

Further, it might be advantageous to implement two different sets of QC samples to the batch sequence, a study QC sample as long-term reference and a batch QC sample as a batch reference [143]. The study QC sample, prepared once for the entire study from a representative sample matrix in advance, enables the inter-batch comparison of data acquisition performance and QC-based post-acquisitional data normalization. The batch QC sample, being a pool of lipid extracts of a distinct batch prepared batch-wise after sample lipid extraction, allow the



determination of batch performance and might also serve as a control of the sample preparation procedure of each batch.

In terms of the LC method, an optimization of the gradient run time might be reasonable to improve the chromatographic separation of triacylglycerides eluting lately in RP separations due to their hydrophobicity. Since the lipid class of TGs is structurally highly diverse, caused by three fatty acyl chains connected to the glycerol backbone, they exhibit a large number of isomeric and isobaric compounds. In the currently applied LC method with a total run time of 15 min (including 3 min of equilibration), the linearity of retention time with carbon chain length and degree of saturation is impaired for TG species eluting between 11.5 min and 12 min. Both lipid identification, i.e. enabling identification on molecular lipid species level, and quantification can be improved by a better chromatographic separation.

Regarding the data processing workflow, optimization potential is given in terms of data extraction, lipid identification, and quantification. Currently, the targeted data extraction is performed on MS1 level by the use of precursor ion masses. By the inclusion of specific fragments on MS2 level to the target list, the optimum quantifier in terms of sensitivity and specificity can be selected [66, 85, 200]. Further, lipid quantification can be optimized by the application of a larger set of IS for a better representation of structural diversity (see above) and establishing a surrogate calibration approach for a more accurate than relative lipid quantification as presented by Drotleff et al. [85]. However, due to the limitations in sample matrix availability and sample pooling of platelets for the generation of matrix-matched calibration samples, this approach has to be developed further for the application in platelet lipidomics.

Finally, the lipid identification rate in the presented lipidomics studies was around 20-25%, thus, the identity of the majority of detected features was unknown. Consequently, there is a high potential to increase the identification rate, but also the identification validity. By the use of additional lipid annotation software (e.g. LipidMatch [48]), different databases (e.g. LMSD [209], HMDB [210], SwissLipids [211]), and the increase of experimentally generated reference spectra based on authentic standards rather than *in-silico* created reference spectra in the lipid databases, both annotation rate and fidelity can be improved [208]. The enhanced translation of analytical responses to lipid information can contribute significantly to the biological and clinical study output.

#### **4.8 Global platelet lipid profiling - Insights into cardiovascular disease, its therapy, and risk assessment**

Large-scale clinical metabolite profiling datamatrices contain a plenty of valuable information regarding clinical or biological questions. Consequently, the same lipidomics data can be analyzed repetitively with varying groups for comparison or statistical tools in dependence on the concrete study interest. The platelet lipid profiling data of the pre- and main study datamatrices were subjected to statistical analysis in the context of different clinical questions.

#### 4.8.1 Significant lipidome alterations in dependence on disease severity

The role of platelets and their activation state is known to be critical in terms of thrombo-inflammation, atheroprogession, and arising adverse cardiovascular events, e.g. myocardial infarction or ischaemic stroke, in patients suffering coronary artery disease (CAD) [36–39]. Based on an increase in uptake and intracellular concentrations of oxidized low-density lipoproteins (oxLDL), the involvement of the platelet lipidome in platelet hyperreactivity in acute coronary syndrome (ACS) was shown recently [43, 212, 213]. To further evaluate lipidome alterations in dependence on disease severity of CAD, the platelet lipid profiles of the prestudy patients cohort were investigated comparing patients suffering acute and chronic coronary syndrome (CCS). The results were published by Harm et al. in 2021 [214] and an overview is presented below.

Among the CAD cohort (baseline characteristics summary in Figure A.1), 28% of the patients exhibited a troponin-positive ACS including ST- and non-ST-segment myocardial infarction. The statistical evaluation was based on univariate, i.e. one-way ANOVA with sequential goodness of fit (SGoF) p-value adjustment to account for multiple hypothesis testing, and multivariate methods like principle component analysis (PCA) and orthogonal partial least squares analysis with discriminant analysis (OPLS-DA). Normalized peak intensities of 928 identified lipids were compared between ACS and CCS patients. The univariate statistics revealed significant lipid alterations ( $P < 0.05$ ,  $FDR < 5\%$ ) of 32 lipids (3% of the identified lipidome) with upregulation of seven and downregulation of 25 lipids in ACS (Figures 4.35 and 4.36).

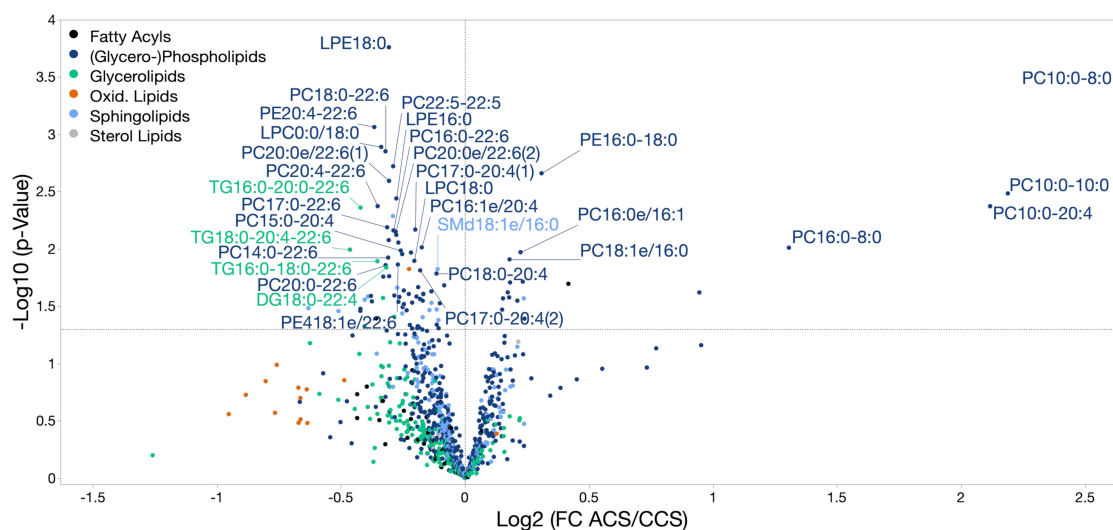


Figure 4.35: Volcano plot of identified lipids. Logarithmic transformed p-values obtained by ANOVA were plotted against the logarithmic (base 2) fold change (FC) in normalized peak intensities of ACS ( $n=26$ ) against CCS ( $n=79$ ), the horizontal line represents the cut-off value of  $P < 0.05$  ( $-\log_{10}(1.3)$ ).  $FC > 1$  indicates an upregulation and  $FC < 1$  downregulation in ACS. Significant features (SGof adjusted  $P < 0.05$ ) are labeled. Reprinted with permission from [214] (Copyright 2021, European Society of Cardiology).

All seven upregulated lipids were glycerophospholipids at least i. being highly saturated, ii. comprising a medium chain fatty acyl side chain, or iii. having an alkyl-/alkenyl bonding to the glycerol backbone. In contrast, downregulated lipids were glycerophospholipids ( $n=20$ ), glycerolipids ( $n=4$ ), and sphingolipids ( $n=1$ ) exhibiting i. a high degree of unsaturation or ii. being deacylated. The lipid identifications of all significantly altered lipids were inspected in detail again to finally confirm the lipid notations. Further, for the medium chain fatty acyl PCs (MCFA-PCs; PC 10:0-8:0, PC 10:0-10:0, PC 10:0-20:4) and one plasmalogen PC (PC-O 16:0/16:1 = PC 16:0e/16:1), additional targeted measurements were performed for side chain clarification, since those species were only identified on lipid species level by MS-DIAL. In addition, upregulated lipids were inspected for isomeric and isobaric overlap in accordance with LSI recommendations [135] (Supplemental Material from Harm et al. [214]). Both additional curation steps confirmed the validity of the lipid species.

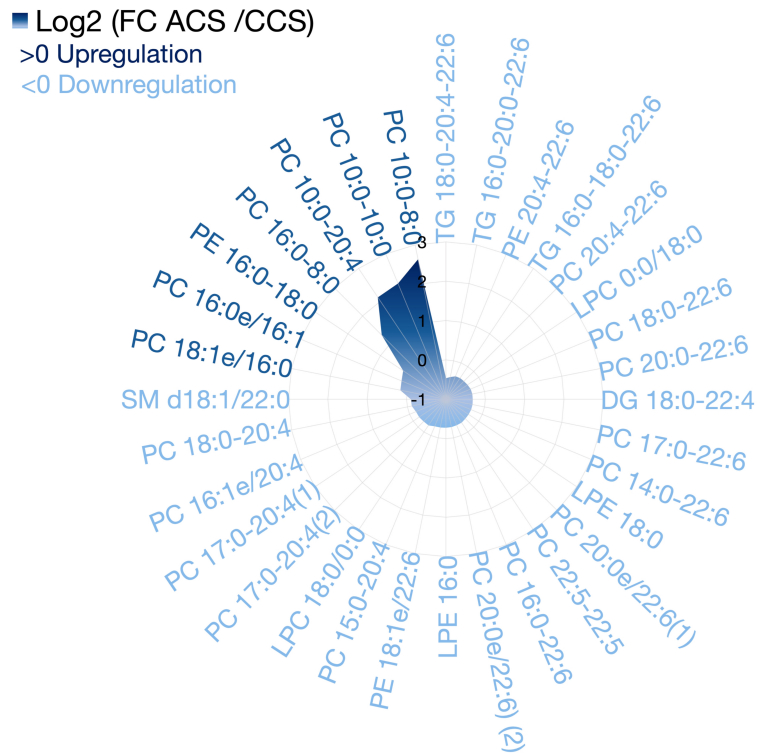


Figure 4.36: Radar chart of significantly altered lipids (SGoF  $P < 0.05$ , FDR  $< 5\%$ ). It shows seven upregulated and 25 downregulated lipids in ACS ( $n=26$ ) compared to CCS patients ( $n=79$ ). Regioisomers, constitutional isomers, or multiple listing due to ionization in both polarity modes are indicated by numbers in brackets. Reprinted with permission from [214] (Copyright 2021, European Society of Cardiology).

The largest lipid level alteration in dependence on disease severity was observed for the MCFA-PCs, being upregulated in the lipidome of ACS patients in an extent of up to a six-fold increase compared to CCS patients. Due to the complexity of the clinical background of this cohort, i.e. comorbidities and comedications, the effect of confounding factors on the disease-state dependent lipidome alterations were assessed by subgroup analysis using multivariate statistical methods (medication with P2Y12-inhibitors and statins, hyperlipidemia and obesity). No confounding factor affected substantially the detected lipid alterations in disease severity comparison [214].

A summary of all significantly altered lipids with concentrations among each clinical group and the results from the statistical analysis is given in Table 4.13.

Table 4.13: Lipid concentrations of significantly up- and downregulated lipids in ACS patients compared to CCS. Concentrations were determined by relative quantification to lipid class specific internal standards and are given as median values with interquartile range (IQR) among each group in pmol/10<sup>6</sup> cells. Reprinted with permission from [214] (Copyright 2021, European Society of Cardiology)

| regulation     | lipid             | ACS [pmol/10 <sup>6</sup> cells]<br>median [IQR] | CCS [pmol/10 <sup>6</sup> cells]<br>median [IQR] | p-value <sup>1</sup> | p-value (SGoF) <sup>1</sup> |
|----------------|-------------------|--------------------------------------------------|--------------------------------------------------|----------------------|-----------------------------|
| upregulation   | PC 10:0-8:0       | 5.61E-04 [3.49E-04 - 6.55E-03]                   | 3.61E-04 [1.97E-04 - 7.33E-04]                   | 0.0004               | 0.0042                      |
|                | PC 10:0/10:0      | 9.91E-04 [2.32E-04 - 6.00E-03]                   | 3.63E-04 [1.80E-04 - 7.62E-04]                   | 0.0033               | 0.0103                      |
|                | PC 10:0-20:4      | 1.12E-03 [3.30E-04 - 8.30E-03]                   | 5.90E-04 [3.43E-04 - 1.43E-03]                   | 0.0042               | 0.0127                      |
|                | PC 16:0-8:0       | 5.40E-03 [2.01E-03 - 1.71E-02]                   | 4.14E-03 [1.93E-03 - 7.50E-03]                   | 0.0097               | 0.0267                      |
|                | PE 16:0-18:0      | 0.35 [0.26 - 0.47]                               | 0.30 [0.21 - 0.41]                               | 0.0022               | 0.0075                      |
|                | PC 16:0e/16:1     | 0.16 [0.14 - 0.20]                               | 0.14 [0.11 - 0.17]                               | 0.0107               | 0.0277                      |
|                | PC 18:1e/16:0     | 3.93 [3.60 - 4.82]                               | 3.72 [3.16 - 4.42]                               | 0.0123               | 0.0286                      |
| downregulation | SM d18:1/22:0     | 47.40 [42.50 - 55.84]                            | 48.70 [41.89 - 55.35]                            | 0.0150               | 0.0341                      |
|                | PC 18:0-20:4      | 19.35 [16.76 - 21.57]                            | 21.95 [17.88 - 24.33]                            | 0.0164               | 0.0405                      |
|                | PC 16:1e/20:4     | 3.60 [2.69 - 4.00]                               | 3.72 [3.23 - 4.48]                               | 0.0097               | 0.0262                      |
|                | PC 17:0-20:4      | 1.05 [0.87 - 1.18]                               | 1.22 [1.00 - 1.52]                               | 0.0068               | 0.0174                      |
|                | PC 15:0-20:4      | 1.32 [1.02 - 1.72]                               | 1.54 [1.22 - 1.90]                               | 0.0103               | 0.0270                      |
|                | PC 18:0-22:6      | 1.67 [1.20 - 2.22]                               | 1.91 [1.50 - 2.56]                               | 0.0088               | 0.0260                      |
|                | PE 18:1e/22:6     | 6.65 [5.08 - 7.61]                               | 6.75 [5.67 - 8.80]                               | 0.0137               | 0.0302                      |
|                | LPE 16:0          | 0.31 [0.27 - 0.36]                               | 0.39 [0.34 - 0.46]                               | 0.0036               | 0.0107                      |
|                | PC 16:0-22:6      | 4.61 [3.31 - 6.27]                               | 5.28 [4.20 - 6.49]                               | 0.0069               | 0.0218                      |
|                | PC 22:5-22:5      | 6.04E-02 [4.32E-02 - 8.97E-02]                   | 7.20E-02 [5.37E-02 - 1.03E-01]                   | 0.0019               | 0.0073                      |
|                | PC 20:0e/22:6     | 0.10 [0.06 - 0.12]                               | 0.11 [0.08 - 0.14]                               | 0.0025               | 0.0088                      |
|                | LPE 18:0          | 1.24 [1.14 - 1.46]                               | 1.52 [1.32 - 1.76]                               | 0.0002               | 0.0036                      |
|                | PC 17:0-22:6      | 5.03E-02 [3.72E-02 - 5.43E-02]                   | 5.74E-02 [4.16E-02 - 7.70E-02]                   | 0.0065               | 0.0155                      |
|                | DG 18:0-22:4      | 1.94E-02 [1.33E-02 - 2.58E-02]                   | 2.44E-02 [1.73E-02 - 3.10E-02]                   | 0.0145               | 0.0328                      |
|                | PC 20:0-22:6      | 9.14E-02 [6.28E-02 - 1.30E-01]                   | 0.12 [0.08 - 0.16]                               | 0.0138               | 0.0309                      |
|                | LPC 18:0          | 3.95 [3.35 - 4.71]                               | 4.65 [3.95 - 5.31]                               | 0.0013               | 0.0068                      |
|                | PC 20:4-22:6      | 1.88 [1.41 - 2.59]                               | 2.39 [1.86 - 3.13]                               | 0.0042               | 0.0111                      |
|                | TG 16:0-18:0-22:6 | 6.64E-02 [4.72E-02 - 8.27E-02]                   | 7.94E-02 [5.81E-02 - 1.11E-01]                   | 0.0128               | 0.0294                      |
|                | PE 20:4-22:6      | 0.52 [0.41 - 0.71]                               | 0.70 [0.53 - 0.85]                               | 0.0009               | 0.0054                      |
|                | TG 16:0-20:0-22:6 | 2.17E-02 [1.63E-02 - 2.96E-02]                   | 3.11E-02 [2.19E-02 - 4.78E-02]                   | 0.0044               | 0.0132                      |
|                | TG 18:0-20:4-22:6 | 3.87E-02 [3.32E-02 - 5.85E-02]                   | 5.63E-02 [3.72E-02 - 8.17E-02]                   | 0.0101               | 0.0270                      |

<sup>1</sup> Statistical analysis based on normalized peak intensities.

Among the statistically significant lipids being altered between the platelet lipidome in ACS and CCS patients, the upregulated lipids were further focused since MCFA-containing lipids are rarely reported in platelets and typically associated with metabolic disorders [215]. The concentration levels in CAD patients were compared to levels in the lipidome of healthy controls and the corresponding boxplots are shown in Figure 4.37.

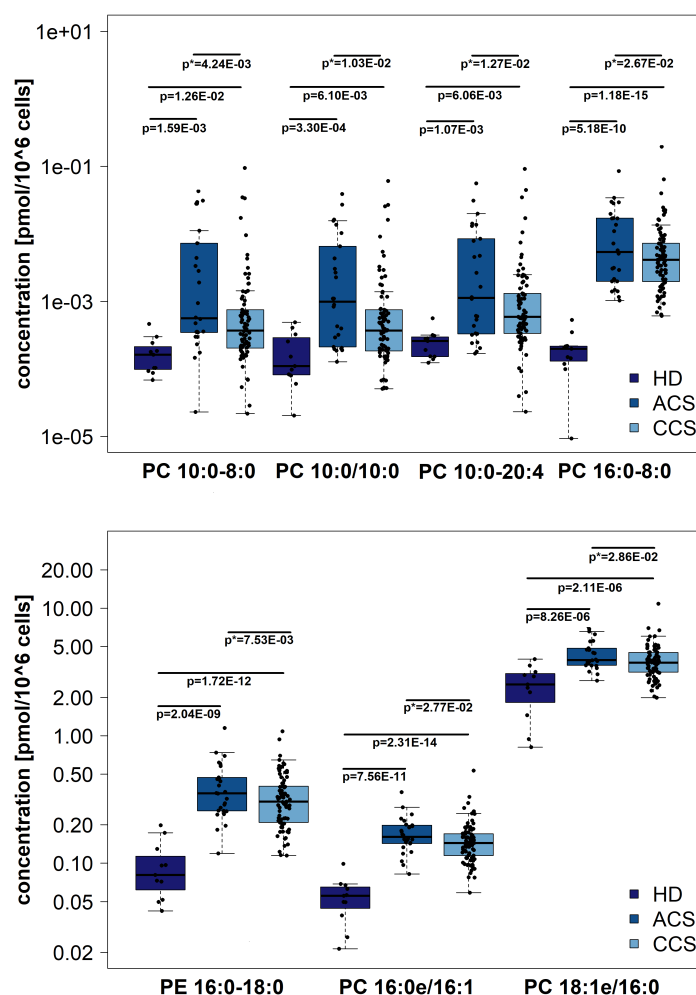
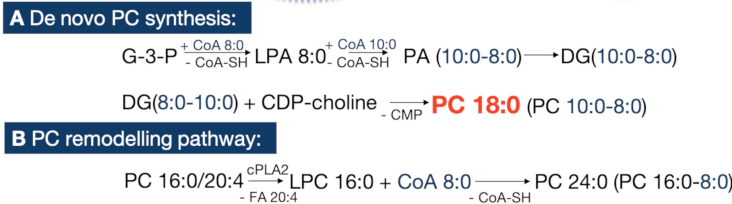


Figure 4.37: Boxplots of significantly upregulated lipids in ACS (n=26) compared to CCS (n=79) and healthy donors (HD, n=11). P-values were obtained by one-way ANOVA and, for ACS-CCS comparison, adjusted by SGoF for multiple hypothesis testing. Thus, adjusted p-values are given for ACS-CCS comparison, p-values were unadjusted for ACS-HD and CCS-HD comparisons. Significance is indicated by \*. Boxplots in (A) show the results for MCFA-PCs. For HD, the lipid concentrations were below the limit of detection (LOD, S/N<3) and peak intensities were obtained by gap filling (data extraction of the local maximum in a defined mass range and retention time window). (B) shows the plots of PE and alkyl-/alkenyl PCs. Reprinted with permission from [214] (Copyright 2021, European Society of Cardiology).

The comparison of the platelet lipid profile of ACS and CCS patients revealed significant differences in the lipidome depending on disease severity. As previously described in literature, platelet reactivity is increased in ACS patients [216, 217] and platelet lipidome alterations modulate the cellular function [212, 213]. Therefore, it was assumed that significantly regulated lipids might contribute to an altered platelet function. Supported by *in vitro* function

assays, especially PC 18:0 seems to have prothrombotic effects causing an increase in *in vitro* platelet-dependent thrombus formation, aggregation and spreading of platelets, and  $\text{Ca}^{2+}$ -signaling [214]. These results are indicative for a pathophysiological role of MCFA-PCs in ACS. Further, due to their essential role as membrane building components, alterations in the levels of glycerophospholipids modulate the platelet function [218, 219] and, based on the lipidomics and experimental *in vitro* data, it was assumed that platelet hyperreactivity is associated with the upregulated PC lipid species, especially PC 18:0 (Figure 4.38).



VLCFA=very long chain FA. Reprinted with permission from [214] (Copyright 2021, European Society of Cardiology).

As a hypothesis, MCFA-PCs might originate from a fatty acid oxidation disorder in peroxisomes and/or in mitochondria. An increase in peroxisomal  $\beta$ -oxidation activity of very long chain fatty acids (VLCFA) to octanoyl-CoA or a decreased further  $\beta$ -oxidation to acetyl-CoA in mitochondria might lead to an accumulation of MCFA-CoA, causing an incorporation into glycerol(phospho)lipids by de novo synthesis or remodelling pathways [214]. Supportingly, increased levels of octanoyl-CoA in CAD patients in contrast to healthy controls were found in a targeted quantification approach for acyl-CoAs [220].



Besides the prothrombotic effect of the upregulated lipids in ACS, the role of the downregulated lipids might be associated with the common structural feature being shared by the most of them, a polyunsaturated fatty acyl (PUFA) side chain. PUFA-containing glycerophospholipids are responsible for membrane integrity and fluidity [221, 222] and are involved in thrombo-inflammatory signaling by releasing the PUFA side chain intracellularly by phospholipase activity (cPLA2) for further oxidation to lipid mediators [223]. Depending on the progenitor PUFA, e.g. arachidonic acid (FA 20:4) or docosahexaenoic acid (FA 22:6), proinflammatory or proresolving mediators are generated. Thus, a dysregulation in PUFA-PL levels might contribute to stronger thrombo-inflammation in ACS [214]. Further, PUFA-PLs might be involved in ferroptotic processes. Due to the phospholipid peroxidation without subsequent reduction to the corresponding alcohol due to GPx4 deficiency, cell death by ferroptosis is induced [224, 225]. PUFA side chains in glycerophospholipids are known as sensitizer for ferroptosis [224], thus a decrease in PUFA-PLs might be a result of an increase of phospholipid peroxidation in ACS patients.

#### 4.8.2 Effects of statin treatment on the platelet lipidome

In addition to the role of the platelet lipidome in the disease severity status, platelet lipid levels might be effected by medications for CAD treatment. An enhanced platelet activation status is associated with an increased risk for thrombo-ischemic events in CAD patients [38, 39, 43]. Since cardiovascular risk factors like dyslipoproteinemia or diabetes contribute to platelet hyperreactivity, medical treatment of these diseases assist the prevention of adverse cardiovascular events in CAD patients [40, 43]. Statins, having a lipid-lowering effect by reducing the synthesis of cholesterol through 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibition, are often used drugs in patients under cardiovascular risk to reduce plasma concentrations of low-density lipoprotein (LDL)-cholesterol [226]. Thus, the effect of statins on the platelet lipidome was investigated in the prestudy cohort of patients suffering symptomatic CAD (baseline characteristics in Figure A.2). The results were published by Harm et al. in 2023 [227] and are summarized below.

Within the CAD cohort, 71 patients (68%) were treated with statins compared to 34 non-treated patients. The identified lipidome (n=928) was investigated for statin-dependent alterations by uni- and multivariate statistical methods. The data analysis revealed a significantly different platelet lipidome between statin-treated and statin-naïve patients (Figure 4.39), showing an upregulation of triacylglyceride (TG) species in statin-treated patients as most decisive alteration with 35 species among the upregulated lipids (n=41; TGs: 85%) and all significantly altered lipids (n=47; TGs: 74%), respectively [227]. Furthermore, cholesteryl esters with PUFAs linked to cholesterol by an ester bond (CE 20:4, CE 20:5, CE 22:4, CE 22:6) were observed among the significantly upregulated lipids in statin-treated patients compared to statin-naïve. Only a few lipid species were downregulated among statin treatment, comprising of glycerophospholipids (PEs: PE 36:4; PCs: PC 16:0/18:3, PC 18:3/16:0, PC 18:2-18:3, PC 18:2-18:2) and sphingolipids (Cer-NDS d18:0/24:0).

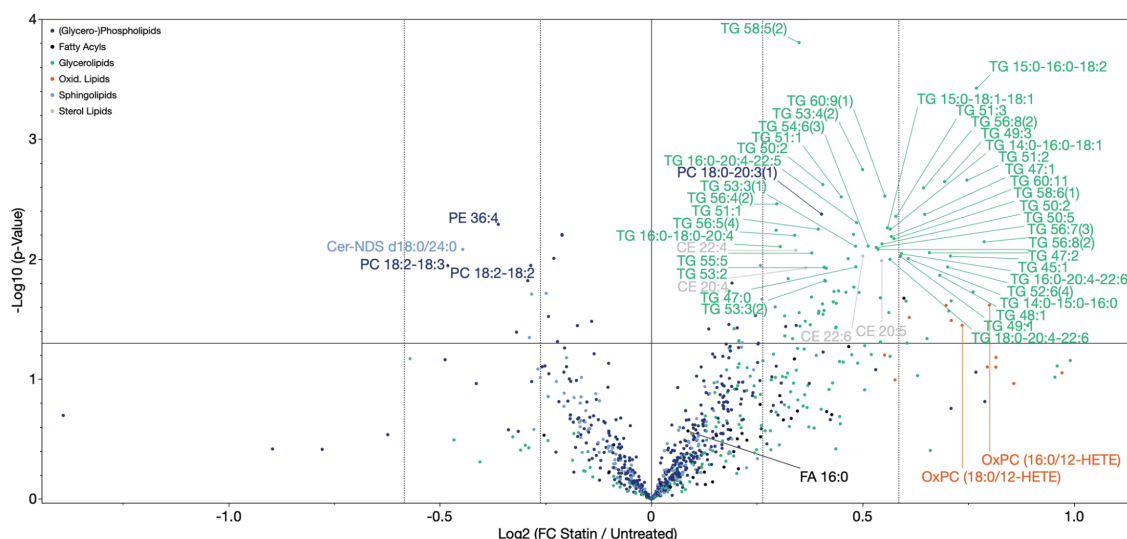


Figure 4.39: Volcano plot illustrating lipidome alterations depending on statin treatment. Normalized peak intensities of 928 identified lipids were compared between statin-treated ( $n=71$ ) and statin-naïve patients ( $n=34$ ) and the  $\log_{10}$ -transformed p-values obtained by one-way ANOVA were plotted against the  $\log_2$ -transformed FC. Here, a  $FC > 1$  represents an upregulation in the statin-treated group,  $FC < 1$  a downregulation, respectively. Significantly altered lipids (SGoF adjusted p-value  $< 0.05$ , FDR  $< 5\%$ ) are labeled by the lipid species name. Furthermore, FA 16:0 and two enzymatically oxidized PCs are labeled. Reprinted with permission from [227] (Copyright 2023, Thieme).

The univariate statistics revealed significant lipid differences in the platelet lipidome in dependence on statin treatment, mostly comprising of species from TG and PC lipid classes. Therefore, a more detailed view into alterations among all detected species of these lipid classes between statin-treated and statin-naïve patients was obtained by respective lipid class heatmaps shown in Figure 4.40. Lipid species are characterized by the total carbon number of the side chains and the total number of double bonds, while their alteration in dependence on the statin treatment and the statistical significance is indicated by the color and size of the dots [227].

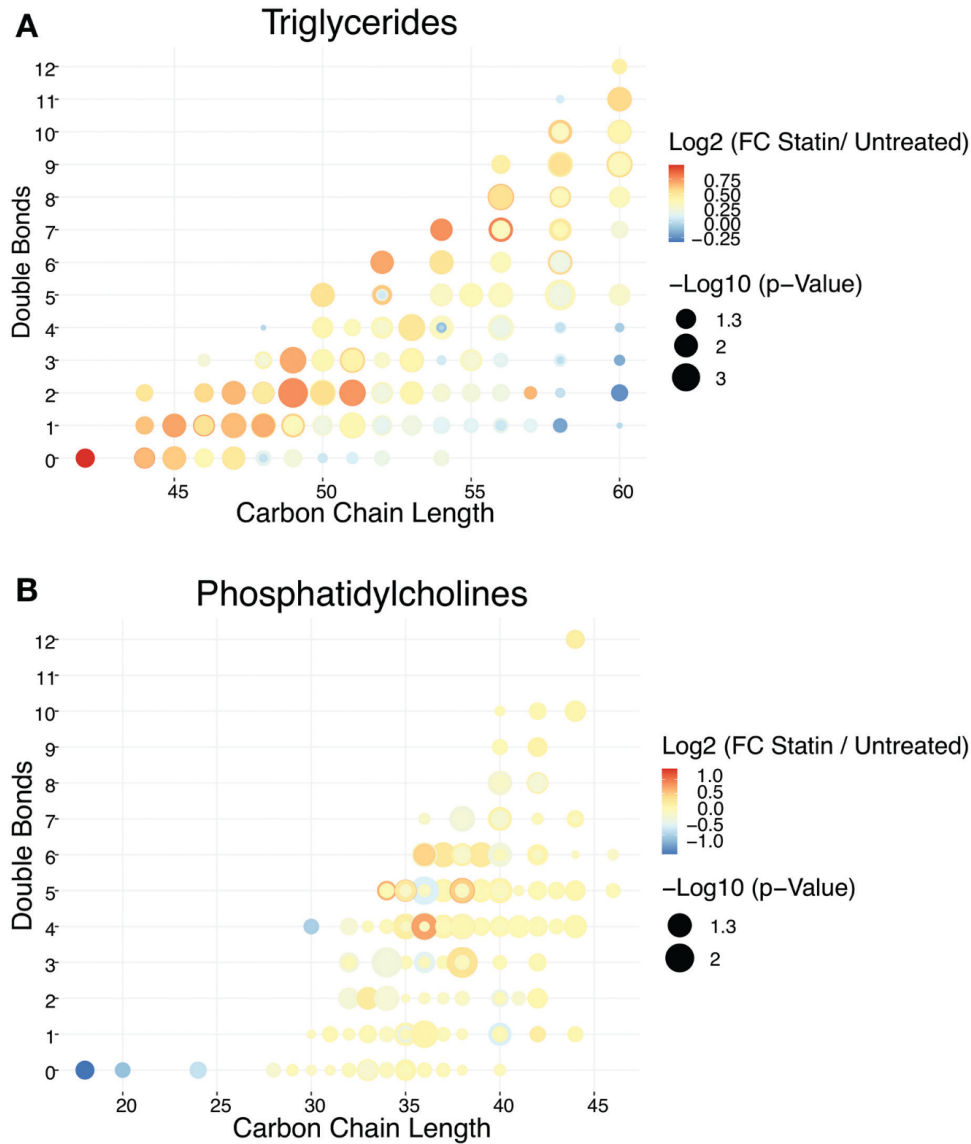


Figure 4.40: Heat maps of TG and PC lipid classes. Within each lipid class, the number of double bonds were plotted against the total side chain carbon number. Each lipid species is represented by a dot, color and size of the dots indicate the logarithmic transformation of the FC and the p-value in the comparison of statin-treated (n=71) and statin-naïve subcohorts (n=34). Reprinted with permission from [227] (Copyright 2023, Thieme).

The TGs heat map indicates the upregulation of the majority of lipid species with the strongest effect in lipids with cumulative side chain carbon numbers of 42 to 51 carbons and a relatively low degree of saturation (cumulative number of DB<3) [227]. Among the PC lipid class, in general less intense and significant alterations were observed. Here, it was striking that the MCFA-PCs, found to be upregulated in ACS patients and related to the disease severity with supposed hyperreactivity promoting effects, were downregulated in the statin-treated cohort compared to statin-naïve patients [227]. It might be assumed that the statin treatment had a

softening effect on the platelet lipidome, contributing to less prothrombotic phenotype [227].

Hence, a disease-dependent statin effect on the platelet lipidome was further investigated by multivariate statistics comparing statin-treated vs. statin-naïve patients in the ACS and CCS subcohorts (Figure 4.41).

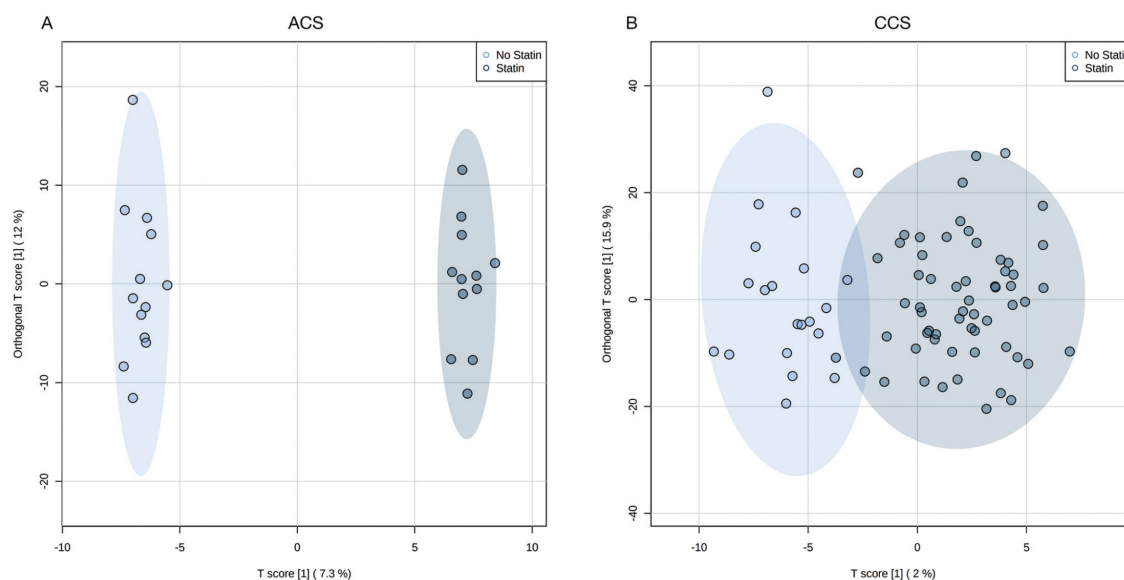


Figure 4.41: Score plot of OPLS-DA analysis of lipidome alterations in dependence on statin-treatment in ACS and CCS subcohorts. The OPLS-DA was based on the identified lipidome with normalized peak intensities. In (A), statin-treated ( $n=12$ ) and statin-naïve patients ( $n=13$ ) in a subcohort of ACS patients is shown. In (B), statin-dependent platelet lipidome alterations in the CCS subcohort are depicted (statin-treated:  $n=20$ , statin-naïve:  $n=60$ ). Reprinted with permission from [227] (Copyright 2023, Thieme).

The OPLS-DA revealed a stronger effect of statin-dependent alterations among the ACS subcohort than in CCS patients, also mainly caused by an upregulation of TG lipid species [227]. Further, downregulated lipids mainly comprised of very long chain (L)PCs, plasmalogen-, or PUFA-PCs. Thus, statin effects on membrane fluidity due to the membrane location of PCs, but also effects on the platelet activation, i.e. modulated by peroxisomal or phospholipase activity, might be assumed [227].

The impact evaluation of statin medication on the platelet lipidome in the prestudy CAD cohort was indicative for a pleiotropic effect of statins on platelets by softening the lipidome [227]. Platelet hyperreactivity is critical in the disease state of CAD, especially in ACS [36, 37, 214, 228].

The main statin-dependent effect on the lipidome was the upregulation of species from the TG lipid class. The majority of the altered TGs contained less unsaturated fatty acyl side chains, i.e. FA 18:1 and FA 18:2. Since TGs constitute the reservoir of fatty acids for building PC lipids, they contribute to platelet integrity and reactivity [227]. Especially PCs with FA 18:2 side chains provide the precursor lipid for 13-HODE which exhibits antithrombotic

properties by inhibition of thromboxane A<sub>2</sub> synthesis [229]. Further, the upregulation of PUFA-containing CE, i.e. storage of PUFAs, leads to the assumption that the generation of PUFA-PLs, having an essential role in membrane integrity and thrombo-inflammation, is reduced. PUFA-PLs provide precursors for both proinflammatory and proresolving lipid mediators, thus, modulate the thrombo-inflammation [227, 230].

Additionally, a downregulation of MCFA-PCs, being correlated with disease severity, was observed in statin-treated compared to statin-naïve patients [227]. MCFA-PCs were assumed to be upregulated in ACS patients due to an impaired mitochondrial and/or peroxisomal  $\beta$ -oxidation leading to an accumulation of MCFA-CoAs [214, 220]. Those lipids were associated with platelet hyperreactivity, increasing the risk of adverse cardiovascular events in CAD patients [214]. The statin-dependent lipidome alterations, i.e. upregulation of TGs, CEs, and FA 16:0 as well as downregulation of MCFA-PCs, might be indicative for a pleiotropic effect of statins on the counteracting pathways of fatty acid synthesis and oxidation in platelets [214, 227]. A reduction of  $\beta$ -oxidation might cause the decrease of MCFA-PCs, an increase of fatty acid synthesis might lead to higher levels of FA 16:0 and, due to fatty acid incorporation for storage, of TGs and CEs [227, 231]. These conditions are usually observed when cells require less energy [231]. Moreover, these effects were shown to be dose-dependent, thus, higher statin doses caused stronger lipid alterations [227]. Consequently, it was hypothesized that statins modulate the platelet lipidome by alterations in the lipid metabolism of fatty acid oxidation and synthesis with beneficial effects on platelet functions in cardiovascular diseases. Due to the antithrombotic and lipidome softening pleiotropic effects, the relevance of statin treatment beyond its lipid-lowering effects was highlighted in this clinical study [227].

#### 4.8.3 Cardiovascular risk assessment in CVD patients

One goal of large-scale prospective clinical studies, as conducted in the main platelet lipidomics study, is the elucidation of distinct lipid species serving as lipid biomarkers for an increased risk of secondary cardiovascular events. Hence, the clinical CAD cohort (n=1057, baseline characteristics in Figure A.3) was evaluated for patients suffering follow-up events (i.e cardiac death, ischemic stroke, myocardial infarction, bleeding; n=72 (6.8%), Figure A.4) within three years after study enrollment, and their platelet lipidomes were compared with eventless patients for significant lipid alterations being associated with cardiovascular risk assessment. The results are summarized below, based on the publication "Large-scale lipidomics profiling reveals characteristic lipid signatures associated with an increased cardiovascular risk" by Harm et al. 2023 [232].

To evaluate lipidome changes, univariate statistics with false discovery rate corrections, i.e. student's t-test followed by Benjamini-Hochberg (BH) procedure was applied for comparing lipid concentrations of 767 identified lipids from CAD patients without any event against patients suffering cardiovascular (cardiac death, ischemic stroke, myocardial infarction, n=51) or bleeding events (n=21). Multivariate analysis by PCA resulted in a clear separation of

patients with events and those without, emphasizing different lipid profiles (Figure 4.42) [232].

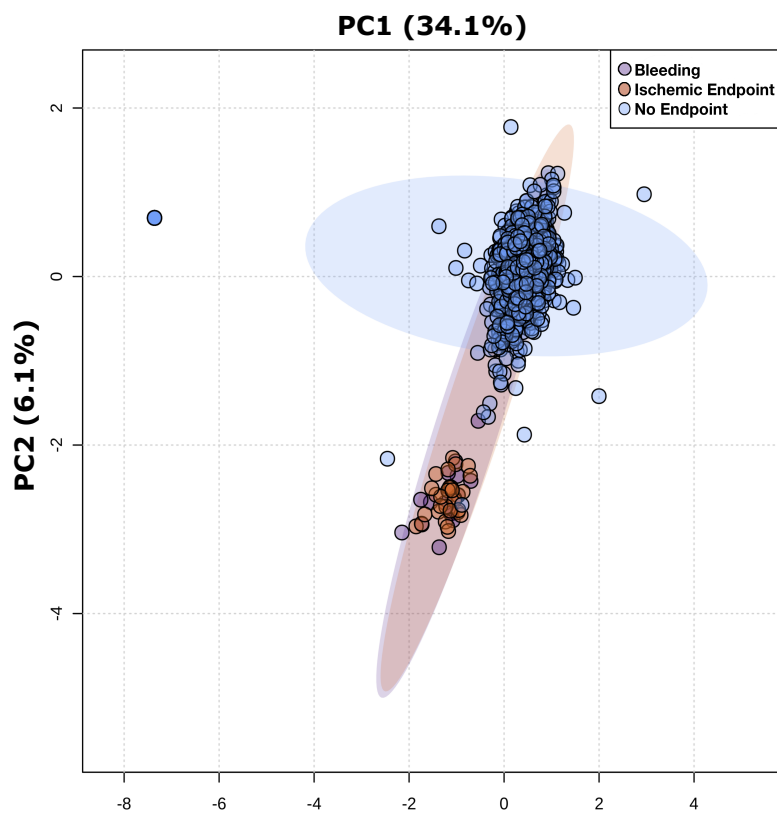


Figure 4.42: PCA of the identified platelet lipidome from the CAD cohort (767 lipids) indicating separation of patients due to ischemic (orange), bleeding (purple), or no cardiovascular events (light blue). Reprinted after modification from [232] (Copyright 2023, the Authors).

As indicated by volcano plots in Figure 4.43 and the summary in Table 4.14, 18 lipids were significantly upregulated ( $p < 0.05$ ,  $\text{FDR} \leq 5\%$ ) in patients with ischemic events, all originating from glycerophospholipid category (Table 4.14). In contrast, platelet lipid profiles of patients with bleeding events prominently exhibited a significant increase in acylcarnitines and fatty acids (Table 4.14) [232].

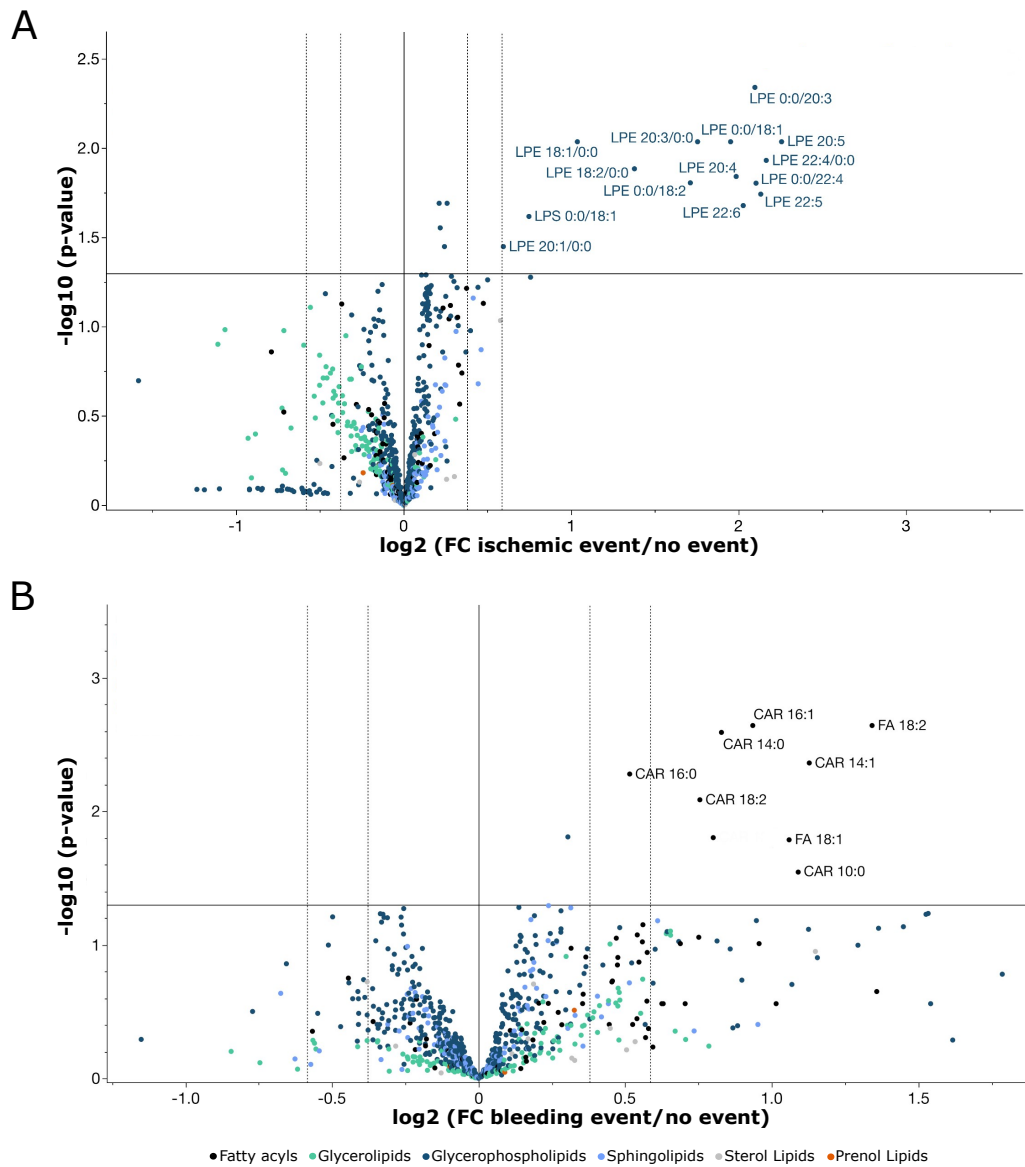


Figure 4.43: Volcano plots differentiating eventless CAD patients from patients suffering adverse (A) ischemic (n=51) or (B) bleeding (n=21) events. The plots were based on lipid concentrations of 767 identified lipids and the log-transformed p-values were plotted against the logarithmic FC, while significantly regulated lipids (BH  $p < 0.05$ ) with a  $FC > 1.5$  were labeled by the respective lipid notation. Reprinted after modification from [232] (Copyright 2023, the Authors).

Table 4.14: Summary of significantly upregulated lipids comparing patients with ischemic (n=51) or bleeding events (n=21) with patients without events (n=985). Results were obtained by student's t-test with BH FDR correction ( $p < 0.05$ ,  $FDR \leq 5\%$ ) based on lipid concentrations of the identified lipidome [232]

| ischemic events | bleeding events |
|-----------------|-----------------|
| LPE 18:1/0:0    | CAR 10:0        |
| LPE 0:0/18:1    | CAR 12:0        |
| LPE 18:2/0:0    | CAR 14:0        |
| LPE 0:0/18:2    | CAR 14:1        |
| LPE 20:1/0:0    | CAR 16:0        |
| LPE 20:3/0:0    | CAR 16:1        |
| LPE 0:0/20:3    | CAR 18:2        |
| LPE 0:0/20:4    | FA 18:1         |
| LPE 0:0/20:5    | FA 18:2         |
| LPE 22:4/0:0    | PC 18:0-22:5    |
| LPE 0:0/22:4    |                 |
| LPE 0:0/22:5    |                 |
| LPE 0:0/22:6    |                 |
| PE 16:0-18:1    |                 |
| PC 18:1-20:1    |                 |
| PC-O 16:0-20:1  |                 |
| PC-O 36:2       |                 |
| LPS 0:0/18:1    |                 |

Further lipidome analysis revealed that both ischemic and bleeding events were associated with larger side chain length and degree of unsaturation in LPE and TG lipid classes while inverse effects were observed for PCs [232].

The potential of the identified lipidome alterations in assessing the risk for adverse ischemic events was evaluated by Cox proportional hazard models being based on the upregulated lyso-glycerophospholipids (Table 4.14). After adjustment for age, gender, and statin medication, all lysoPLs exhibited a hazard ratio larger one independently, emphasizing an association of these lipids with an increased cardiovascular risk (Table 4.15). Additionally, independent risk prediction for bleeding events revealed an association with FA 18:1 and FA 18:2 (Table 4.15) [232].



Table 4.15: Hazard ratios (HR) from Cox proportional hazard models indicating the independent association of significantly upregulated lipids with the cardiovascular risk of ischemic or bleeding events. Data were adjusted for age, gender, and statin therapy. HR calculation with 95% confidence interval was performed for 1-SD higher lipid concentrations [232]

| <i>Ischemic events</i> |                   |
|------------------------|-------------------|
| <b>lipid species</b>   | <b>HR (95%CI)</b> |
| LPE 20:1/0:0           | 1.18 (1.07-1.29)* |
| LPS 0:0/18:1           | 1.16 (1.04-1.28)* |
| LPE 0:0/20:3           | 1.14 (1.06-1.23)* |
| LPE 20:3/0:0           | 1.14 (1.06-1.23)* |
| LPE 18:1/0:0           | 1.14 (1.06-1.23)* |
| LPE 0:0/18:1           | 1.14 (1.06-1.23)* |
| LPE 0:0/20:4           | 1.14 (1.05-1.23)* |
| LPE 18:2/0:0           | 1.14 (1.05-1.23)* |
| LPE 0:0/22:4           | 1.13 (1.05-1.22)* |
| LPE 0:0/18:2           | 1.13 (1.05-1.22)* |
| LPE 22:4/0:0           | 1.13 (1.05-1.22)* |
| LPE 0:0/22:6           | 1.13 (1.05-1.22)* |
| LPE 0:0/20:5           | 1.13 (1.05-1.22)* |
| LPE 0:0/22:5           | 1.13 (1.05-1.22)* |
| <i>Bleeding events</i> |                   |
| <b>lipid species</b>   | <b>HR (95%CI)</b> |
| FA 18:1                | 1.16 (1.00-1.40)* |
| FA 18:2                | 1.14 (1.00-1.32)* |
| CAR 14:1               | 1.03 (0.91-1.18)  |
| CAR 16:1               | 1.02 (0.88-1.19)  |
| CAR 14:0               | 1.01 (0.85-1.22)  |
| CAR 10:0               | 1.01 (0.85-1.21)  |
| CAR 16:0               | 1.01 (0.83-1.22)  |

\* p-value<0.05.

The association of LPE and acylcarnitine lipids with increased risks of adverse ischemic or bleeding events was further supported by *ex vivo* function assays of platelets at the time of blood sampling [Harm et al. 2023, submitted]. Since platelet aggregation was correlated with increasing LPE concentrations and decreasing acylcarnitine levels, it was assumed that LPEs contribute to enhanced platelet activation while acylcarnitines are involved in lowering the hyperreactivity of platelets in CVD patients [232].

Lysophosphatidylethanolamines are known to contribute to platelet hyperreactivity [233], and based on the findings of this study, promote platelet aggregation and increase the risk for adverse cardiovascular events in CAD [Harm et al. 2023, submitted]. Regarding a higher risk of bleeding events in CAD patients, the increase in fatty acid and acylcarnitine levels seems to have inhibitory effects on platelet function. The association of FA and ACars with the severe course of CAD was already found in plasma [212, 234]. By lowering the platelet activity and potentiating antiplatelet therapy [235–237] as well as exhibiting anticoagulative

effects [238], fatty acids and acylcarnitines, respectively, modulate platelet function [232].

For the evaluation of potential biomarkers in risk stratification in CVD, cross-validated LASSO selection algorithm was applied to all identified lipids after adjustment for age, gender, and statin treatment. The analysis revealed the relation of nine lipids with adverse cardiovascular events, i.e. PC 34:2;O, LPE 20:3/0:0, PI 18:1-20:4, PE-O 16:1-18:2, PI 16:0-20:4, FA 18:2;2O, FA 18:1, TG 14:0-16:0-18:1, TG 16:0-14:1-18:1 [Harm et al. 2023, submitted]. These lipids were added in stepwise models to conventional risk factors, i.e. age, gender, diabetes mellitus, smoking, arterial hypertension, total cholesterol, HDL-cholesterol, to assess the three-year risk of developing cardiovascular events. The prediction, indicated by the C index, was improved by the addition of the selected lipids up to a C index of 0.739 (95% CI 0.703-0.816) compared to C index 0.647 (95% CI 0.612-0.741) of the conventional model. Further, including the lipid panel by replacing cholesterol measures (total and HDL-cholesterol) also resulted in an enhanced predictive value (C index 0.741, 95% CI 0.695-0.812). This supports the value of including platelet lipids to conventional risk markers in cardiovascular risk assessment [232].

The platelet lipidome is critically involved in the pathophysiology of CAD, contributing to hyperreactivity [36, 37, 228], more severe disease status [214], and higher risk of adverse cardiovascular events, but its therapeutic modulation might lead to a beneficial course of disease [227, 239]. Consequently, distinct platelet lipids might have the potential for diagnostic improvements or paving the way to new therapeutic strategies.

#### 4.8.4 Platelet-derived therapeutic targets - ACKR3/CXCR7

Besides the stand-alone evaluations of the large-scale clinical lipidomics data as presented above, these datasets can be further used to complement and support experimental studies of platelet-related thrombotic and thrombo-inflammatory mechanisms. Therefore, lipid profiles of CAD patients from the prestudy cohort were integrated in investigations on the platelet surface receptor ACKR3/CXCR7 and its modulation of thrombotic and thrombo-inflammatory processes in CAD patients [240]. The chemokine receptors ACKR3/CXCR7 and CXCR4 regulate platelet response after chemokine ligation of CXCL12/SDF1 $\alpha$  [241], resulting in enhanced platelet activation, aggregation, and cell death through CXCR4-CXCL12/SDF1 $\alpha$  interaction or in cell survival by ACKR/CXCR7-CXCL12/SDF1 $\alpha$  ligation [22]. Due to the prosurvival modulation abilities of the ACKR3/CXCR7 receptor, it was evaluated as therapeutic target in cardiovascular disease [242].

In the study published by Cebo et al. [240], the effects of ACKR3/CXCR7 agonist (VUF11207) administration on the platelet lipidome with consequences for the thrombotic and thrombo-inflammatory response were investigated. A reduction in platelet aggregation was observed for CAD patients exhibiting higher levels of ACKR3/CXCR7 receptor on the cell surface. Administration of ACKR3/CXCR7 agonist led to a decrease in prothrombotic platelet activity *ex vivo* and a reduction of thrombotic and thrombo-inflammatory processes without

affecting physiological haemostasis *in vivo* [240]. Based on lipidomics analysis, performed by targeted analysis of platelet lipid mediators and by untargeted lipid profiling in experimental samples of CAD patients, higher levels of lipids with antithrombotic effect, i.e. dihomo- $\gamma$ -linolenic acid (DGLA), 12-hydroxyeicosatrienoic acid (12-HETrE), compared to atherogenic and prothrombotic lipids were induced by agonist ligation with ACKR3/CXCR7 receptors [240]. The lipidomics data from the prestudy CAD cohort were investigated for lipidome alterations in dependence on antiplatelet therapy (APT) to assess the effects of conventional treatment strategies compared to the potential of ACKR3/CXCR7 therapeutic opportunities. In detail, lipid class levels of DG, LPC, LPI were compared in patients with and without APT, further differences in lipid classes under thrombin activation and thrombin activation accompanied with ACKR3/CXCR7 agonist treatment were evaluated compared to resting status (Figure 4.44).

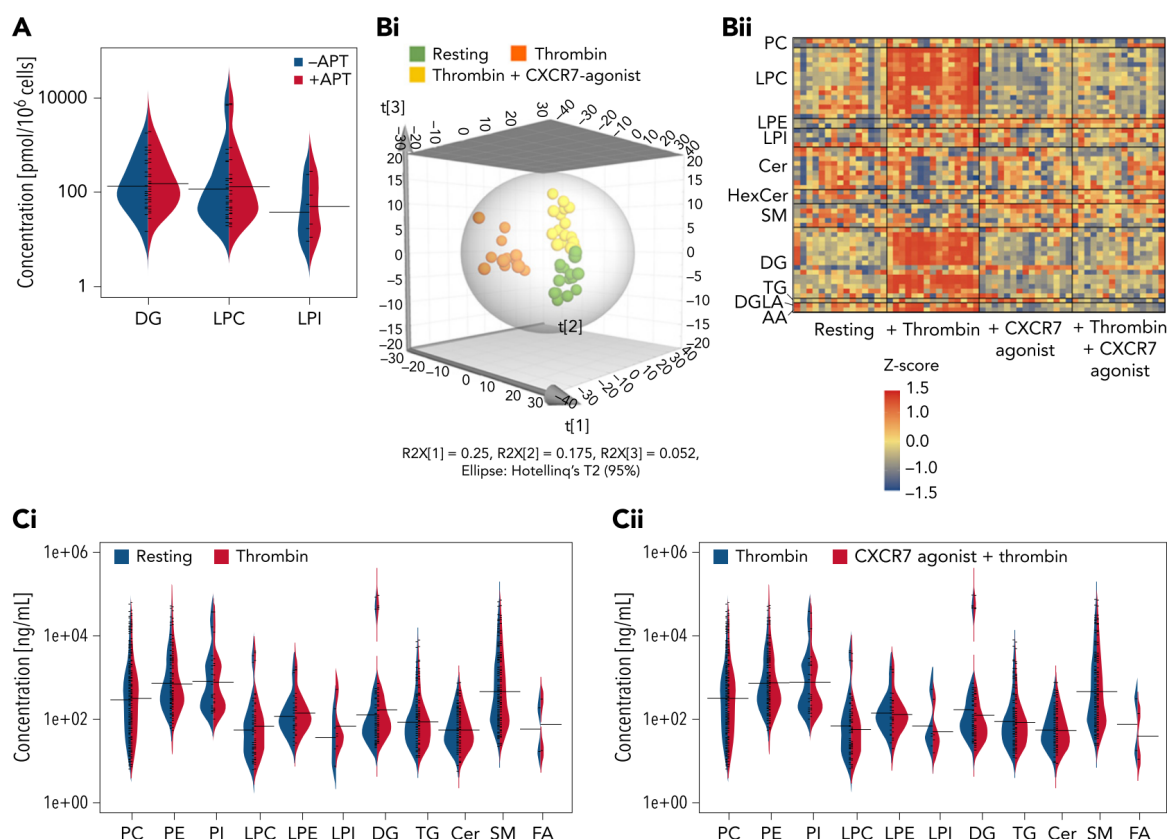


Figure 4.44: Evaluation of platelet lipidome alterations in dependence on APT in the clinical prestudy cohort and effects of *ex vivo* ACKR3/CXCR7 agonist treatment of CAD platelet samples. Data are based on lipid concentrations obtained from relative quantification with lipid class specific internal standards. In (A), asymmetric beanplots show lipid concentration distributions in DG, LPC, LPI lipid classes in CAD patients with APT ( $n=46$ ) and without APT ( $n=61$ ). Here, large lines represent the mean concentration in each lipid class and subgroup, while small lines indicate lipid species concentrations in each subgroup. The plots from (B) and (C) are derived from experimental *ex vivo* data of platelets from CAD patients ( $n=15$ ) in resting status, after thrombin-induced activation (0.1 U/mL for 15 min), and CXCR7 agonist treatment (100  $\mu\text{g/mL}$ ) prior thrombin activation. The PLS-DA model (Bi) shows a clear separation of sample groups due to their activation and treatment status (six components;  $R^2X(\text{cum})=0.566$ ,  $R^2Y(\text{cum})=0.971$ , and  $Q^2(\text{cum})=0.780$ ), which is also indicated by the heatmap (Bii). The asymmetric beanplots show different lipid levels in distinct lipid classes between resting and by thrombin activated platelets (Ci) and platelet lipidome after thrombin-induced activation in absence or presence of CXCR7 agonist treatment (Cii). Reproduced with permission from [240] (Copyright 2022, American Society of Hematology).

The evaluation of prothrombotic, i.e. DGs and LPIs, and atherogenic, i.e. LPCs, lipids in the clinical CAD cohort revealed no significant lipid alterations in dependence on APT, emphasizing the need of further therapeutic strategies to reduce thrombosis and thrombo-inflammation in CAD [240]. In contrast, CXCR7 agonist ligation with ACKR3/CXCR7 receptor counteracted the thrombin-induced lipidome alterations, amongst others DG, LPC, and LPI lipid

classes, thus, exhibiting an antithrombotic effect [240]. Hence, the clinical lipidomics data of the prestudy CAD cohort contributed valuable information to the assessment of platelet lipidome alterations upon therapeutic treatments, highlighting the benefits of CXCR7 agonist over conventional APT by acetylsalicylic acid in antithrombotic therapies.

## 5 Conclusion

In the recent years, the relevance of lipidomics in bioanalytical studies increased significantly and also in clinical research, lipid profiling emerged to an important technique to gain insights in physiological and pathophysiological processes, to identify new therapeutic targets, or to characterize lipid biomarkers for diagnostic, prognostic, or therapeutic purposes. In this thesis, clinical lipidomics studies on the platelet lipidome of coronary artery disease patients were conducted to elucidate lipidome alterations in dependence on different clinical conditions with the perspective of biomarker development. The lipidome analysis contributed new aspects of both analytical and clinical relevance. From the analytical viewpoint, lipid analysis was performed in a profiling, untargeted approach using SWATH for the comprehensive acquisition of MS1 and MS2 data, which was newly applied to a large-scale clinical cohort of more than 1000 subjects. Large-scale metabolomics studies of such size are often, when performed in untargeted approaches, conducted on MS1 level (HUSERMET metabolic profiling cohort[132], ADNI1 Alzheimer disease cohort [243]), or MS2 data are acquired in a data dependent manner only on a limited set of pooled samples (GOLDN cohort [244]), or targeted data analysis is applied (LIPID and ADVANCE CVD cohorts [245, 246]). Consequently, a new processing workflow combining untargeted with targeted data extraction was established. Regarding the clinical aspect, lipidomics in CVD research is often based on plasma or serum samples, thus the presented studies were novel in the investigation of the lipidome of platelets in a large clinical CAD cohort.

From a limited, but still large-scale (>100 samples) patients cohort in the prestudy setup, an analytical workflow had to be established since the existing data processing procedure for untargeted lipidomics was impractical for the application in multi-batch large-scale measurements. Thus, a combined untargeted-targeted data extraction strategy was successfully developed being based on the generation of a reference target list for targeted data extraction originating from untargeted data extraction combined with inter-batch feature alignment. Targeted extraction of untargeted acquired data is already established in proteomics [126] and lipidomics [66, 200], offering a straightforward data processing even in large-scale applications regarding computational capacities and data curation efforts being required in automated peak picking and feature annotation procedures in untargeted data processing. However, the latter procedure enables unsupervised feature extraction without the prerequisite of defining lipid targets in advance, hence unexpected new lipid features, even being of unknown identity initially, might be discovered. In the prestudy, MCFA-PCs were identified as highly relevant lipid alterations being associated with disease severity. Those lipids were not covered in targeted data extraction procedures in lipidomics [66, 200], thus the new workflow was crucial for the clinical findings in the prestudy cohort. Through further workflow refinement during the main study data processing, the quality of the clinical lipidomics datamatrix was improved by increased data curation efforts, causing a reduction of targets in the reference list by misannotated lipids as well as irreproducible and artefact signals (22% of target entries).

Although the generation of the reference target list for the platelet lipidome was time- and labor-intensive, data curation is mandatory for high quality lipidomics data and confident study conclusions. Furthermore, the created reference list is representative for the human platelet lipidome, thus, can be universally applied to other platelet lipidomics studies without the need of rerunning the entire processing workflow.

A disadvantage of untargeted data extraction is caused by differently detected lipidomes in different studies. The inter-study comparability can be increased when applying targeted data extraction with identical target lists. However, to compare the platelet lipidome from different study cohorts post-processing, the inter-batch feature alignment tool, being developed during the workflow optimization of the prestudy processing, offers the opportunity of inter-study lipidome alignment. Thus, terminated lipidomics studies might be reused to support new investigations, e.g. serving for cross-validation evaluations or providing additional samples like healthy donors. The reusability is one of the four cornerstones of the FAIR (Findability, Accessibility, Interoperability, Reuse) principles, which gain importance in scientific research.

Although the lipidome of platelets is highly regulated in dependence on the activation state and alterations are associated with disease progression or medical treatment in CAD patients, lipid profile evaluations of platelets in large clinical CAD cohorts are lacking so far. Thus, with the large-scale clinical lipidomics studies in this thesis, the gap was closed. Lipid profiling data enable the generation of hypotheses on different clinical conditions and the identification of potential biomarkers. The observed lipid alterations in dependence on disease severity and statin treatment were indicative for substantial lipid metabolism alterations, postulated to involve peroxisomal and mitochondrial  $\beta$ -oxidation and metabolic changes caused by oxidative stress and hypoxic conditions (Figure 5.1) [227].

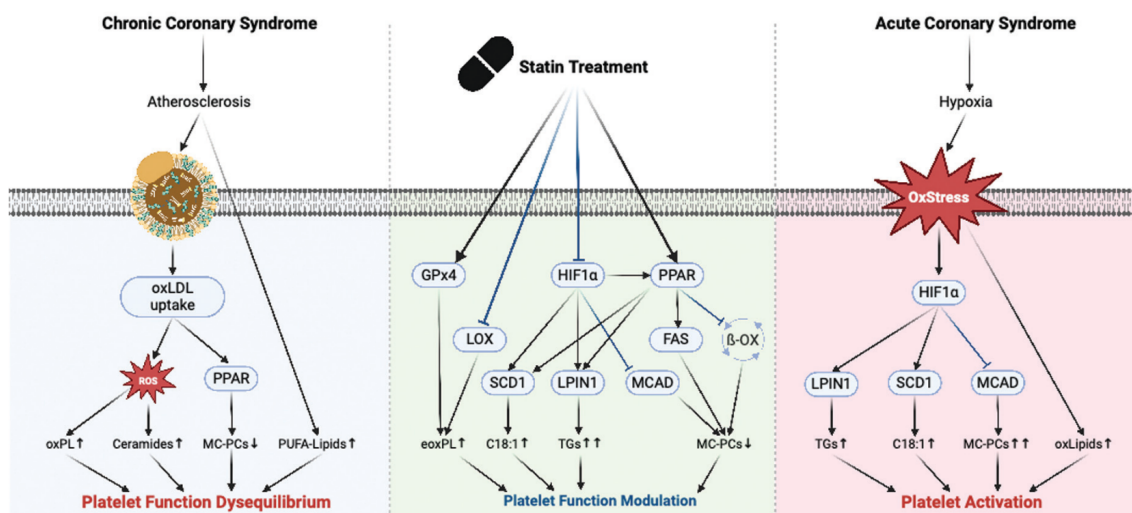


Figure 5.1: Postulated metabolic pathway alterations in chronic and acute coronary syndrome and the effect of statin medication on the platelet lipid metabolism. Abbreviations:  $\beta$ OX= $\beta$ -oxidation, FAS=fatty acid synthetase, GPx4=glutathione peroxidase, HIF1 $\alpha$ =hypoxia-inducible factor 1  $\alpha$ , LOX=lipoxygenase, LPIN1=phosphatidate phosphatase, MC-PC=medium-chain phospholipids, oxLDL=oxidized low-density lipoprotein, PPAR=peroxisome proliferator-activated receptor, ROS=reactive oxygen species, SCD1=stearoyl-CoA desaturase. Reproduced with permission from [227] (Copyright 2023, Thieme).

To conclude, the analytical development of the post-acquisitional data processing strategy for large-scale untargeted lipidomics led to the preparation of highly informative lipid profiling data of clinical CAD patients cohorts and enabled the elucidation of lipid alterations and metabolic pathway mechanisms in different clinical conditions. The results may serve as the basis for further biomarker investigation to improve diagnostic or prognostic assessments or highlight new therapeutic strategies. Furthermore, the potential of the lipid datamatrices is not ladled yet, hence, they may build the base for future analytical and clinical research.



## References

- (1) Van der Meijden, P. E. J.; Heemskerk, J. W. M. *Nat Rev Cardiol* **2019**, *16*, 166–179.
- (2) Jackson, S. P. *Nat Med* **2011**, *17*, 1423–36.
- (3) Gawaz, M.; Langer, H.; May, A. E. *J Clin Invest* **2005**, *115*, 3378–84.
- (4) Semple, J. W.; Italiano J. E., J.; Freedman, J. *Nat Rev Immunol* **2011**, *11*, 264–74.
- (5) Thomas, S. G. In *Platelets (Fourth Edition)*, Michelson, A. D., Ed., Fourth Edition; Academic Press: 2019, pp 47–77.
- (6) Bishop-Bailey, D. *Steroids* **2010**, *75*, 570–5.
- (7) Quach, M. E.; Chen, W.; Li, R. *Blood* **2018**, *131*, 1512–1521.
- (8) Hla, T.; Galvani, S.; Rafii, S.; Nachman, R. *Journal of Experimental Medicine* **2012**, *209*, 2137–2140.
- (9) Broos, K.; De Meyer, S. F.; Feys, H. B.; Vanhoorelbeke, K.; Deckmyn, H. *Thromb Res* **2012**, *129*, 245–9.
- (10) O'Donnell, V. B.; Murphy, R. C.; Watson, S. P. *Circ Res* **2014**, *114*, 1185–203.
- (11) Walther, T.; Chung, J.; Farese Jr., R. *Annual Review of Cell and Developmental Biology* **2017**, *33*, 491–510.
- (12) Fitch-Tewfik, J.; Flaumenhaft, R. *Frontiers in Endocrinology* **2013**, *4*, DOI: 10.3389/fendo.2013.00077.
- (13) Nurden, A. T. *Thromb Haemost* **2011**, *105*, S13–S33.
- (14) Neumüller, J.; Ellinger, A.; Wagner, T. In *The Transmission Electron Microscope*, Maaz, K., Ed.; IntechOpen: Rijeka, 2015; Chapter 11.
- (15) Ponomareva, A.; Nevzorova, T.; Mordakhanova, E.; Andrianova, I.; Litvinov, R. *Cell Tiss. Biol.* **2016**, *10*, 217–226.
- (16) Rumbaut, R.; Thiagarajan, P. *Morgan & Claypool Life Sciences* **2010**.
- (17) Selvadurai, M. V.; Hamilton, J. R. *Platelets* **2018**, *29*, PMID: 29442528, 319–325.
- (18) Bye, A. P.; Unsworth, A. J.; Gibbins, J. M. *J Thromb Haemost* **2016**, *14*, 918–30.
- (19) Weyrich, A. S.; Schwertz, H.; Kraiss, L. W.; Zimmerman, G. A. *J Thromb Haemost* **2009**, *7*, 241–6.
- (20) Miller, V. M.; Jayachandran, M.; Hashimoto, K.; Heit, J. A.; Owen, W. G. *Gend Med* **2008**, *5 Suppl A*, S91–S102.
- (21) Kaushansky, K. *Blood* **2008**, *111*, 981–6.
- (22) Chatterjee M. and Rath, D.; Gawaz, M. *Biochem Soc Trans* **2015**, *43*, 720–6.
- (23) Chatterjee, M. *Cells* **2022**, *11*, DOI: 10.3390/cells11020213.

- (24) WHO <https://www.who.int/publications/i/item/9789240051157> **2022**.
- (25) *BMJ* **2002**, *324*, 71–86.
- (26) Ross, R. *N Engl J Med.* **1999**, *340*, 115–26.
- (27) Furie, B.; Furie, B. *N Engl J Med.* **2008**, *359*, 938–49.
- (28) Burke, A.; Kolodgie, F.; Farb, A.; Weber, D.; Malcom, G.; Smialek, J.; Virmani, R. *Circulation* **2001**, *103*, 934–940.
- (29) Davì, G.; Patrono, C. *N Engl J Med.* **2007**, *357*, 2482–94.
- (30) Davies, M. *Circulation* **1990**, *82*, 1138–46.
- (31) Mann, J.; Davies, M. *Heart* **1999**, *82*, 265–268.
- (32) Lusis, A. *Nature* **2000**, *407*, 233–41.
- (33) Spickett, C.; Dever, G. *Biofactors* **2005**, *24*, 17–31.
- (34) Libby, P. *Nature* **2002**, *420*, 868–874.
- (35) Trip, M.; Cats, V.; van Capelle, F.; Vreken, J. *N Engl J Med* **1990**, *322*, 1549–54.
- (36) Droppa, M.; Tschernow, D.; Müller, K.; Tavlaki, E.; Karathanos, A.; Stimpfle, F.; Schaeffeler, E.; Schwab, M.; Tolios, A.; Siller-Matula, J.; Gawaz, M.; Geisler, T. *PLoS One* **2015**, DOI: 10.1371/journal.pone.0121620.
- (37) Geisler, T.; Zürn, C.; Simonenko, R.; Rapin, M.; Kraibooj, H.; Kilias, A.; Bigalke, B.; Stellos, K.; Schwab, M.; May, A.; Herdeg, C.; Gawaz, M. *Eur Heart J* **2010**, *31*, 59–66.
- (38) Reny, J. et al. *Thromb Haemost* **2016**, *115*, 844–855.
- (39) May, A.; Geisler, T.; Gawaz, M. *Thromb Haemost* **2008**, *99*, 487–93.
- (40) Vinik, A.; Erbas, T.; Park, T.; Nolan, R.; Pittenger, G. *Diabetes Care* **2001**, *24*, 1476–85.
- (41) Rathbone, R.; Ardlie, N.; Schwartz, C. *Pathology* **1970**, *2*, 307–316.
- (42) Kjeldsen, S.; Rostrup, M.; Gjesdal, K.; Eide, I. *Am Heart J* **1991**, *122*, 330–336.
- (43) Stellos, K.; Sauter, R.; Fahrleitner, M.; Grimm, J.; Stakos, D.; Emschermann, F.; Panagiota, V.; Gnerlich, S.; Perk, A.; Schönberger, T.; Bigalke, B.; Langer, H.; Gawaz, M. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2012**, *32*, 2017–2020.
- (44) Fahy, E.; Cotter, D.; Sud, M.; Subramaniam, S. *Biochimica et Biophysica Acta* **2011**, *1811*, 637–647.
- (45) Fahy, E. et al. *J Lipid Res* **2005**, *46*, 839–61.
- (46) Liebisch, G.; Vizcaino, J. A.; Kofeler, H.; Trotzmüller, M.; Griffiths, W. J.; Schmitz, G.; Spener, F.; Wakelam, M. J. *J Lipid Res* **2013**, *54*, 1523–30.
- (47) Liebisch, G. et al. *J Lipid Res* **2020**, *61*, 1539–1555.

- 
- (48) Koelmel, J. P.; Kroeger, N. M.; Ulmer, C. Z.; Bowden, J. A.; Patterson, R. E.; Cochran, J. A.; Beecher, C. W. W.; Garrett, T. J.; Yost, R. A. *BMC Bioinformatics* **2017**, *18*, 331.
- (49) Han, X. *Nature Reviews Endocrinology* **2016**, *12*, 668–679.
- (50) Yeung, J.; Hawley, M.; Holinstat, M. *J Mol Med (Berl)* **2017**, *95*, 575–588.
- (51) Paul, B. Z.; Jin, J.; Kunapuli, S. P. *J Biol Chem* **1999**, *274*, 29108–14.
- (52) Niki, E.; Yoshida, Y.; Saito, Y.; Noguchi, N. *Biochem Biophys Res Commun* **2005**, *338*, 668–76.
- (53) Han, X., *Lipidomics: Comprehensive Mass Spectrometry of Lipids*; Wiley: 2016.
- (54) Nelson, D. L.; Cox, M. M., *Lehninger principles of biochemistry*, 6th; WH Freeman: 2013.
- (55) Yang, K.; Han, X. *Trends Biochem Sci* **2016**, *41*, 954–969.
- (56) Pechlaner, R.; Kiechl, S.; Mayr, M. *Circulation* **2016**, *134*, 1651–1654.
- (57) Lagarde, M.; G  lo  n, A.; Record, M.; Vance, D.; Spener, F. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2003**, *1634*, 61.
- (58) Holcapek, M.; Liebisch, G.; Ekroos, K. *Anal Chem* **2018**, *90*, 4249–4257.
- (59) Tumanov, S.; Kamphorst, J. J. *Curr Opin Biotechnol* **2017**, *43*, 127–133.
- (60) Folch, J.; Lees, M.; Sloane Stanley, G. H. *Journal of Biological Chemistry* **1957**, *226*, 497–509.
- (61) Bligh, E. G.; Dyer, W. J. *Canadian Journal of Biochemistry and Physiology* **1959**, *37*, 911–917.
- (62) Hara, A.; Radin, N. S. *Analytical Biochemistry* **1978**, *90*, 420–426.
- (63) Matyash, V.; Liebisch, G.; Kurzchalia, T. V.; Shevchenko, A.; Schwudke, D. *J Lipid Res* **2008**, *49*, 1137–46.
- (64) Lofgren, L.; Stahlman, M.; Forsberg, G. B.; Saarinen, S.; Nilsson, R.; Hansson, G. I. *J Lipid Res* **2012**, *53*, 1690–700.
- (65) Jurowski, K.; Kochan, K.; Walczak, J.; Bara  nska, M.; Piekoszewski, W.; Buszewski, B. *TrAC Trends in Analytical Chemistry* **2017**, *86*, 276–289.
- (66) Calderon, C.; Rubarth, L.; Cebo, M.; Merfort, I.; Lammerhofer, M. *Proteomics* **2019**, e1900113.
- (67) Sarafian, M. H.; Gaudin, M.; Lewis, M. R.; Martin, F. P.; Holmes, E.; Nicholson, J. K.; Dumas, M. E. *Anal Chem* **2014**, *86*, 5766–74.
- (68) Rustam, Y. H.; Reid, G. E. *Anal Chem* **2018**, *90*, 374–397.
- (69) Wu, Y.; Li, L. *J Chromatogr A* **2016**, *1430*, 80–95.

- 
- (70) Lou, S.; Balluff, B.; Cleven, A.; Bovee, J.; McDonnell, L. *Am Soc Mass Spectrom* **2017**, *28*, 376–383.
- (71) Wang, M.; Wang, C.; Han, R.; Han, X. *Progress in Lipid Research* **2016**, *61*, 83–108.
- (72) Lisa, M.; Cifkova, E.; Khalikova, M.; Ovcacikova, M.; Holčapek, M. *J Chromatogr A* **2017**, *1525*, 96–108.
- (73) Trufelli, H.; Palma, P.; Famiglini, G.; Cappiello, A. *Mass Spectrom Rev* **2011**, *30*, 491–509.
- (74) Gathungu, R.; Larrea, P.; Sniatynski, M.; Marur, V.; Bowden, J.; Koelmel, J.; Starke-Reed, P.; Hubbard, V.; Kristal, B. *Analytical Chemistry* **2018**, *90*, 13523–13532.
- (75) Yang, K.; Cheng, H.; Gross, R.; Han, X. *Analytical Chemistry* **2009**, *81*, 4356–68.
- (76) Quehenberger, O.; Armando, A.; Dennis, E. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2011**, *1811*, 648–656.
- (77) Cajka, T.; Fiehn, O. *Trends Analyt Chem* **2014**, *61*, 192–206.
- (78) Lange, M.; Ni, Z.; Criscuolo, A.; Fedorova, M. *Chromatographia* **2019**, *82*, 77–100.
- (79) Witting, M.; Maier, T.; Garvis, S.; Schmitt-Kopplin, P. *Journal of Chromatography A* **2014**, *1359*, 91–9.
- (80) Bird, S.; Marur, V.; Stavrovskaya, I.; Kristal, B. *Analytical Chemistry* **2012**, *84*, 5509–17.
- (81) Narváez-Rivas, M.; Zhang, Q. *Journal of Chromatography A* **2016**, *1440*, 123–134.
- (82) Cajka, T.; Davis, R.; Austin, K. J.; Newman, J. W.; German, J. B.; Fiehn, O.; Smilowitz, J. T. *Metabolomics* **2016**, *12*, DOI: 10.1007/s11306-016-1075-9.
- (83) Cebo, M.; Fu, X.; Gawaz, M.; Chatterjee, M.; Lammerhofer, M. *J Pharm Biomed Anal* **2020**, *189*, 113426.
- (84) Ovcacikova, M.; Lisa, M.; Cifkova, E.; Holcapek, M. *J Chromatogr A* **2016**, *1450*, 76–85.
- (85) Drotleff, B.; Illison, J.; Schlotterbeck, J.; Lukowski, R.; Lammerhofer, M. *Anal Chim Acta* **2019**, *1086*, 90–102.
- (86) Wang, M.; Wang, C.; Han, X. *Mass Spectrom Rev* **2017**, *36*, 693–714.
- (87) Anesi, A.; Guella, G. *Journal of Chromatography A* **2015**, *1384*, 44–52.
- (88) Líska, M.; Holčapek, M. *Analytical Chemistry* **2015**, *87*, 7187–95.
- (89) Jiang, P.; Lucy, C. *Analytical Methods* **2016**, *8*, 6478–6488.
- (90) Hartmann, A.; Ganzera, M. *Planta Med* **2015**, *81*, 1570–1581.
- (91) Laboureur, L.; Ollero M. Touboul, D. *International Journal of Molecular Sciences* **2015**, *16*, 13868–13884.

- 
- (92) Gross, J., *Mass Spectrometry*, 2nd ed.; Springer Berlin, Heidelberg: 2011.
- (93) Brügger, B. *Annu Rev Biochem* **2014**, *83*, 79–98.
- (94) De Hoffmann, E.; Stroobant, V., *Mass Spectrometry: Principles and Applications*, 3rd ed.; Wiley: 2013.
- (95) Banerjee, S.; Mazumdar, S. *Int J Anal Chem* **2012**, DOI: 10.1155/2012/282574.
- (96) Kool, J.; Niessen, W., *Analyzing Biomolecular Interactions by Mass Spectrometry*; Wiley-VCH: 2015.
- (97) Tang, L.; Kebarle, P. *Anal Chem* **1993**, *65*, 3654–3668.
- (98) Cappiello, A.; Famiglioni, G.; Palma, P.; Pierini, E.; Termopoli V. and Trufelli, H. *Anal Chem* **2008**, *80*, 9343–9348.
- (99) Matuszewski, B. K.; Constanzer, M. L.; Chavez-Eng, C. M. *Analytical Chemistry* **1998**, *70*, 882–889.
- (100) Xian, F.; Hendrickson, C. L.; Marshall, A. G. *Anal Chem* **2012**, *84*, 708–19.
- (101) Hart-Smith, G.; Blanksby, S. J. In *Mass Spectrometry in Polymer Chemistry*; Wiley: 2011; Chapter 1, pp 5–32.
- (102) Urban, P.; Chen, Y.; Wang, Y., *Time-Resolved Mass Spectrometry: From Concept to Applications*; Wiley: 2016.
- (103) Andrews, G. L.; Simons, B. L.; Young, J. B.; Hawkridge, A. M.; Muddiman, D. C. *Anal Chem* **2011**, *83*, 5442–6.
- (104) Miller, P. E.; Denton, M. B. *Journal of Chemical Education* **1986**, *63*, 617.
- (105) Henchman, M.; Steel, C. *Journal of Chemical Education* **1998**, *75*, 1049.
- (106) Gross, J. H. In *Mass Spectrometry: A Textbook*; Springer International Publishing: Cham, 2017, pp 151–292.
- (107) Glush, G. L.; Burinsky, D. J. *Journal of the American Society for Mass Spectrometry* **2008**, *19*, PMID: 18187337, 161–172.
- (108) Somogyi, Á., *Medical Applications of Mass Spectrometry*; Vekey, K., Telekes, A., Vertes, A., Eds.; Elsevier: 2008.
- (109) Johnson, A. R.; Carlson, E. E. *Analytical Chemistry* **2015**, *87*, PMID: 26132379, 10668–10678.
- (110) Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. *J Mass Spectrom* **2001**, *36*, 849–65.
- (111) Collins, R. *Vacuum* **1969**, *19*, 105–111.
- (112) Chernushevich, I. V. *European Journal of Mass Spectrometry* **2000**, *6*, 471–479.
- (113) Retra, K.; Bleijerveld, O. B.; van Gestel, R. A.; Tielens, A. G.; van Hellemond, J. J.; Brouwers, J. F. *Rapid Communications in Mass Spectrometry* **2008**, *22*, 1853–1862.

- 
- (114) Chao, H. C.; Chen, G. Y.; Hsu, L. C.; Liao, H. W.; Yang, S. Y.; Wang, S. Y.; Li, Y. L.; Tang, S. C.; Tseng, Y. J.; Kuo, C. H. *Anal Chim Acta* **2017**, *971*, 68–77.
- (115) Drotleff, B.; Hallschmid, M.; Lämmerhofer, M. *Analytica Chimica Acta* **2018**, *1022*, 70–80.
- (116) Fillatre, Y.; Rondeau, D.; Jadas-Hécart, A.; Communal, P. Y. *Rapid Communications in Mass Spectrometry* **2010**, *24*, 2453–2461.
- (117) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. *Nat Methods* **2015**, *12*, 523–6.
- (118) Marshall, A. G.; Hendrickson, C. L. *Annual Review of Analytical Chemistry* **2008**, *1*, PMID: 20636090, 579–599.
- (119) Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. *J Am Soc Mass Spectrom* **2016**, *27*, 1897–1905.
- (120) Decaestecker, T.; Clauwaert, K.; Van Bocxlaer, J.; Lambert, W.; Vanden Eeckhout, E.; Van Peteghem, C.; De Leenheer, A. *Rapid Commun Mass Spectrom* **2000**, *14*, 1787–1792.
- (121) Fitzgerald, R.; Rivera, J.; Herold, D. *Clin Chem* **1999**, *45*, 1224–34.
- (122) Hopfgartner, G.; Tonoli, D.; Varesio, E. *Anal Bioanal Chem* **2012**, *402*, 2587–96.
- (123) Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K. *Rapid Communications in Mass Spectrometry* **2006**, *20*, 1989–1994.
- (124) Xu, G.; Stupak, J.; Yang, L.; Hu, L.; Guo, B.; Li, J. *Rapid Communications in Mass Spectrometry* **2018**, *32*, 763–774.
- (125) Simons, B.; Kauhanen, D.; Sylvanne, T.; Tarasov, K.; Duchoslav, E.; Ekroos, K. *Metabolites* **2012**, *2*, 195–213.
- (126) Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. *Mol Cell Proteomics* **2012**, *11*, O111 016717.
- (127) Zhang, Y.; Bilbao, A.; Bruderer, T.; Luban, J.; Strambio-De-Castillia, C.; Lisacek, F.; Hopfgartner, G.; Varesio, E. *J Proteome Res* **2015**, *14*, 4359–71.
- (128) Collins, B. C. et al. *Nat Commun* **2017**, *8*, 291.
- (129) Alvarez-Ruiz, R.; Pico, Y. *Journal of Chromatography A* **2019**, *1595*, 81–90.
- (130) Schlotterbeck, J.; Chatterjee, M.; Gawaz, M.; Lämmerhofer, M. *Anal Chim Acta* **2019**, *1046*, 1–15.
- (131) Raetz, M.; Bonner, R.; Hopfgartner, G. *Metabolomics* **2020**, *16*, 71.

- 
- (132) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R.; Human Serum Metabolome, C. *Nat Protoc* **2011**, *6*, 1060–83.
- (133) Zelena, E.; Dunn, W. B.; Broadhurst, D.; Francis-McIntyre, S.; Carroll, K. M.; Begley, P.; O'Hagan, S.; Knowles, J. D.; Halsall, A.; Consortium, H.; Wilson, I. D.; Kell, D. B. *Anal Chem* **2009**, *81*, 1357–64.
- (134) Dunn, W. B.; Wilson, I. D.; Nicholls, A. W.; Broadhurst, D. *Bioanalysis* **2012**, *4*, 2249–64.
- (135) LSI Lipidomics Standard Initiative. Researcher Consortium, <https://lipidomicstandards.org/>.
- (136) Hyotylainen, T.; Oresic, M. *Anal Bioanal Chem* **2015**, *407*, 4973–93.
- (137) Zivkovic, A. M.; Wiest, M. M.; Nguyen, U. T.; Davis, R.; Watkins, S. M.; German, J. B. *Metabolomics* **2009**, *5*, 507–516.
- (138) Meikle, T. G.; Huynh, K.; Giles, C.; Meikle, P. J. *J Lipid Res* **2021**, *62*, 100127.
- (139) Fu, X.; Calderon, C.; Harm, T.; Gawaz, M.; Lämmerhofer, M. *Analytica Chimica Acta* **2022**, *1221*, 340155.
- (140) Briscoe, C. J.; Stiles, M. R.; Hage, D. S. *J Pharm Biomed Anal* **2007**, *44*, 484–91.
- (141) Wilkinson, M. D. et al. *Sci Data* **2016**, *3*, 160018.
- (142) Antonio, M. G.; Schick-Makaroff, K.; Doiron, J. M.; Shields, L.; White, L.; Molzahn, A. *Western Journal of Nursing Research* **2020**, *42*, PMID: 31665999, 640–648.
- (143) Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B. *Metabolomics* **2018**, *14*, 72.
- (144) Sangster, T.; Major, H.; Plumb, R.; Wilson, A. J.; Wilson, I. D. *Analyst* **2006**, *131*, 1075–8.
- (145) Kamleh, M. A.; Ebbels, T. M.; Spagou, K.; Masson, P.; Want, E. J. *Anal Chem* **2012**, *84*, 2670–7.
- (146) FDA <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioanalytical-method-validation-guidance-industry> **2018**.
- (147) Bruijnsvoort, M. V.; Meijer, J.; den Beld, C. V. *Bioanalysis* **2017**, *9*, 1955–1965.
- (148) Smith, R.; Mathis, A. D.; Ventura, D.; Prince, J. T. *BMC Bioinformatics* **2014**, *15 Suppl 7*, S9.
- (149) Tsugawa, H.; Kanazawa, M.; Ogiwara, A.; Arita, M. *Bioinformatics* **2014**, *30*, 2379–80.
- (150) Pedrioli, P. G. et al. *Nat Biotechnol* **2004**, *22*, 1459–66.

- 
- (151) Patterson, R. E.; Kirpich, A. S.; Koelmel, J. P.; Kalavalapalli, S.; Morse, A. M.; Cusi, K.; Sunny, N. E.; McIntyre, L. M.; Garrett, T. J.; Yost, R. A. *Metabolomics* **2017**, *13*, DOI: 10.1007/s11306-017-1280-1.
- (152) Katajamaa, M.; Oresic, M. *J Chromatogr A* **2007**, *1158*, 318–28.
- (153) Wei, R.; Wang, J.; Su, M.; Jia, E.; Chen, S.; Chen, T.; Ni, Y. *Sci Rep* **2018**, *8*, 663.
- (154) Faquih, T.; van Smeden, M.; Luo, J.; le Cessie, S.; Kastenmuller, G.; Krumsiek, J.; Noordam, R.; van Heemst, D.; Rosendaal, F. R.; van Hylckama Vlieg, A.; Willems van Dijk, K.; Mook-Kanamori, D. O. *Metabolites* **2020**, *10*, DOI: 10.3390/metabo10120486.
- (155) Do, K. T. et al. *Metabolomics* **2018**, *14*, 128.
- (156) Lange, M.; Fedorova, M. *Anal Bioanal Chem* **2020**, *412*, 3573–3584.
- (157) Cajka, T.; Fiehn, O. In *Lipidomics: Methods and Protocols*, Bhattacharya, S. K., Ed.; Springer New York: New York, NY, 2017, pp 149–170.
- (158) De Livera, A. M.; Sysi-Aho, M.; Jacob, L.; Gagnon-Bartsch, J. A.; Castillo, S.; Simpson, J. A.; Speed, T. P. *Anal Chem* **2015**, *87*, 3606–15.
- (159) Fan, S.; Kind, T.; Cajka, T.; Hazen, S. L.; Tang, W. H. W.; Kaddurah-Daouk, R.; Irvin, M. R.; Arnett, D. K.; Barupal, D. K.; Fiehn, O. *Anal Chem* **2019**, DOI: 10.1021/acs.analchem.8b05592.
- (160) Lange, M.; Angelidou, G.; Ni, Z.; Criscuolo, A.; Schiller, J.; Bluher, M.; Fedorova, M. *Cell Rep Med* **2021**, *2*, 100407.
- (161) Ni, Z. et al. *Nat Methods* **2022**, DOI: 10.1038/s41592-022-01710-0.
- (162) Goracci, L.; Tortorella, S.; Tiberi, P.; Pellegrino, R. M.; Di Veroli, A.; Valeri, A.; Cruciani, G. *Anal Chem* **2017**, *89*, 6257–6264.
- (163) Sciex LipidView software. Sciex, <https://sciex.com/products/software/lipidview-software>.
- (164) Song, H.; Ladenson, J.; Turk, J. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009**, *877*, 2847–54.
- (165) Alcoriza-Balaguer, M. I.; Garcia-Canaveras, J. C.; Ripoll-Esteve, F. J.; Lahoz, A. *bioRxiv* **2022**, DOI: 10.1101/2022.02.25.476005.
- (166) Alka, O.; Shanthamoorthy, P.; Witting, M.; Kleigrew, K.; Kohlbacher, O.; Rost, H. L. *Nat Commun* **2022**, *13*, 1347.
- (167) Yin, Y.; Wang, R.; Cai, Y.; Wang, Z.; Zhu, Z.-J. *Analytical Chemistry* **2019**, *91*, PMID: 31436405, 11897–11904.
- (168) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. *Environ Sci Technol* **2014**, *48*, 2097–8.
- (169) Sumner, L. W. et al. *Metabolomics* **2007**, *3*, 211–221.



- 
- (170) Kind, T.; Liu, K. H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O. *Nat Methods* **2013**, *10*, 755–8.
- (171) Blazenovic, I.; Kind, T.; Ji, J.; Fiehn, O. *Metabolites* **2018**, *8*, DOI: 10.3390/metabo8020031.
- (172) De Livera, A. M.; Dias, D. A.; De Souza, D.; Rupasinghe, T.; Pyke, J.; Tull, D.; Roessner, U.; McConville, M.; Speed, T. P. *Anal Chem* **2012**, *84*, 10768–76.
- (173) Luo, P.; Yin, P.; Zhang, W.; Zhou, L.; Lu, X.; Lin, X.; Xu, G. *J Chromatogr A* **2016**, *1437*, 127–136.
- (174) Cairns, D. A.; Thompson, D.; Perkins, D. N.; Stanley, A. J.; Selby, P. J.; Banks, R. E. *PROTEOMICS* **2008**, *8*, 21–27.
- (175) Borrego, S. L.; Fahrman, J.; Datta, R.; Stringari, C.; Grapov, D.; Zeller, M.; Chen, Y.; Wang, P.; Baldi, P.; Gratton, E.; Fiehn, O.; Kaiser, P. *Cancer Metab* **2016**, *4*, 9.
- (176) Gagnebin, Y.; Tonoli, D.; Lescuyer, P.; Ponte, B.; de Seigneux, S.; Martin, P. Y.; Schappler, J.; Bocard, J.; Rudaz, S. *Anal Chim Acta* **2017**, *955*, 27–35.
- (177) Sysi-Aho, M.; Katajamaa, M.; Yetukuri, L.; Orešič, M. *BMC Bioinformatics* **2007**, *8*, DOI: 10.1186/1471-2105-8-93.
- (178) Van der Kloet, F. M.; Bobeldijk, I.; Verheij, E. R.; Jellema, R. H. *J Proteome Res* **2009**, *8*, 5132–41.
- (179) Bijlsma, S.; Bobeldijk, I.; Verheij, E. R.; Ramaker, R.; Kochhar, S.; Macdonald, I. A.; van Ommen, B.; Smilde, A. K. *Anal Chem* **2006**, *78*, 567–74.
- (180) Drotleff, B.; Lammerhofer, M. *Anal Chem* **2019**, *91*, 9836–9843.
- (181) Boysen, A. K.; Heal, K. R.; Carlson, L. T.; Ingalls, A. E. *Anal Chem* **2018**, *90*, 1363–1369.
- (182) Redestig, H.; Fukushima, A.; Stenlund, H.; Moritz, T.; Arita, M.; Saito, K.; Kusano, M. *Anal Chem* **2009**, *81*, 7974–80.
- (183) Li, B.; Tang, J.; Yang, Q.; Cui, X.; Li, S.; Chen, S.; Cao, Q.; Xue, W.; Chen, N.; Zhu, F. *Scientific Reports* **2016**, *6*, DOI: 10.1038/srep38881.
- (184) Li, B.; Tang, J.; Yang, Q.; Li, S.; Cui, X.; Li, Y.; Chen, Y.; Xue, W.; Li, X.; Zhu, F. *Nucleic Acids Res* **2017**, *45*, W162–W170.
- (185) Zacharias, H. U.; Altenbuchinger, M.; Gronwald, W. *Metabolites* **2018**, *8*, DOI: 10.3390/metabo8030047.
- (186) Cleveland, W. S.; J., D. S. *Journal of American Statistical Association* **1988**, *83*, 596–610.
- (187) Luan, H.; Ji, F.; Chen, Y.; Cai, Z. *Anal Chim Acta* **2018**, *1036*, 66–72.

- 
- (188) Wang, L.; Su, B.; Zeng, Z.; Li, C.; Zhao, X.; Lv, W.; Xuan, Q.; Ouyang, Y.; Zhou, L.; Yin, P.; Peng, X.; Lu, X.; Lin, X.; Xu, G. *Analytical Chemistry* **2018**, *90*, 11401–11408.
- (189) Xia, J.; Psychogios, N.; Young, N.; Wishart, D. S. *Nucleic Acids Res* **2009**, *37*, W652–60.
- (190) Chatterjee, M.; Borst, O.; Walker, B.; Fotinos, A.; Vogel, S.; Seizer, P.; Mack, A.; Alampour-Rajabi, S.; Rath, D.; Geisler, T.; Lang, F.; Langer, H. F.; Bernhagen, J.; Gawaz, M. *Circ Res* **2014**, *115*, 939–49.
- (191) Calderon, C.; Sanwald, C.; Schlotterbeck, J.; Drotleff, B.; Lammerhofer, M. *Anal Chim Acta* **2019**, *1048*, 66–74.
- (192) Matuszewski, B. K.; Constanzer, M. L.; Chavez-Eng, C. M. *Analytical Chemistry* **2003**, *75*, 3019–3030.
- (193) Guo, J.; Shen, S.; Xing, S.; Chen, Y.; Chen, F.; Porter, E. M.; Yu, H.; Huan, T. *Anal Chem* **2021**, *93*, 12181–12186.
- (194) Stekhoven, D. J.; Buhlmann, P. *Bioinformatics* **2012**, *28*, 112–8.
- (195) R Core Team R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, <https://www.R-project.org/>.
- (196) Gromski, P. S.; Xu, Y.; Kotze, H. L.; Correa, E.; Ellis, D. I.; Armitage, E. G.; Turner, M. L.; Goodacre, R. *Metabolites* **2014**, *4*, 433–52.
- (197) De Livera, A. M.; Bowne, J. B. Package metabolomics for R. <https://CRAN.R-project.org/src/contrib/Archive/%20metabolomics..>
- (198) De Livera, A. Package MetNorm for R. <https://CRAN.R-project.org/package=MetNorm>.
- (199) De Livera, A. M.; Olshansky, G. Package NormalizeMets for R. <https://CRAN.R-project.org/package=NormalizeMets..>
- (200) Cebo, M.; Calderon Castro, C.; Schlotterbeck, J.; Gawaz, M.; Chatterjee, M.; Lammerhofer, M. *J Pharm Biomed Anal* **2021**, *205*, 114301.
- (201) Cebo, M.; Schlotterbeck, J.; Gawaz, M.; Chatterjee, M.; Lammerhofer, M. *Anal Chim Acta* **2020**, *1094*, 57–69.
- (202) Lange, E.; Tautenhahn, R.; Neumann, S.; Gropl, C. *BMC Bioinformatics* **2008**, *9*, 375.
- (203) Bonner, R.; Hopfgartner, G. *TrAC Trends in Analytical Chemistry* **2018**, DOI: 10.1016/j.trac.2018.10.014.
- (204) Koelmel, J. P.; Li, X.; Stow, S. M.; Sartain, M. J.; Murali, A.; Kemperman, R.; Tsugawa, H.; Takahashi, M.; Vasiliou, V.; Bowden, J. A.; Yost, R. A.; Garrett, T. J.; Kitagawa, N. *Metabolites* **2020**, *10*, DOI: 10.3390/metabo10030101.

- (205) Hartler, J. et al. *Nat Methods* **2017**, *14*, 1171–1174.
- (206) McDonald, J. G.; Ejsing, C. S.; Kopczynski, D.; Holčapek, M.; Aoki, J.; Arita, M.; Arita, M.; Baker, E. S.; Bertrand-Michel, J.; Bowden, J. A. *Nature Metabolism* **2022**, *4*, 1086–1088.
- (207) Evans, A. M. et al. *Metabolomics* **2020**, *16*, 113.
- (208) Theodoridis, G.; Gika, H.; Raftery, D.; Goodacre, R.; Plumb, R. S.; Wilson, I. D. *Anal Chem* **2023**, *95*, 3909–3916.
- (209) Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill Alfred H., J.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Subramaniam, S. *Nucleic Acids Research* **2006**, *35*, D527–D532.
- (210) Wishart, D. S. et al. *Nucleic Acids Research* **2012**, *41*, D801–D807.
- (211) Aimo, L.; Liechti, R.; Hyka-Nouspikel, N.; Niknejad, A.; Gleizes, A.; Götz, L.; Kuznetsov, D.; David, F. P.; van der Goot, F. G.; Riezman, H.; Bougueleret, L.; Xenarios, I.; Bridge, A. *Bioinformatics* **2015**, *31*, 2860–2866.
- (212) Chatterjee, M.; Rath, D.; Schlotterbeck, J.; Rheinlaender, J.; Walker-Allgaier, B.; Alnaggar, N.; Zdanyte, M.; Muller, I.; Borst, O.; Geisler, T.; Schaffer, T. E.; Lammerhofer, M.; Gawaz, M. *Eur Heart J* **2017**, *38*, 1993–2005.
- (213) Peng, B.; Geue, S.; Coman, C.; Munzer, P.; Kopczynski, D.; Has, C.; Hoffmann, N.; Manke, M. C.; Lang, F.; Sickmann, A.; Gawaz, M.; Borst, O.; Ahrends, R. *Blood* **2018**, DOI: 10.1182/blood-2017-12-822890.
- (214) Harm, T.; Bild, A.; Dittrich, K.; Goldschmied, A.; Nestele, J.; Chatterjee, M.; Fu, X.; Kolb, K.; Castor, T.; Borst, O.; Geisler, T.; Rath, D.; Lammerhofer, M.; Gawaz, M. *Cardiovasc Res* **2021**, DOI: 10.1093/cvr/cvab238.
- (215) Vishwanath, V. A. *Ann Neurosci* **2016**, *23*, 51–5.
- (216) Fuchs, I.; Frossard, M.; Spiel, A.; Riedmuller, E.; Laggner, A. N.; Jilma, B. *J Thromb Haemost* **2006**, *4*, 2547–52.
- (217) Frossard, M.; Fuchs, I.; Leitner, J. M.; Hsieh, K.; Vlcek, M.; Losert, H.; Domanovits, H.; Schreiber, W.; Laggner, A. N.; Jilma, B. *Circulation* **2004**, *110*, 1392–1397.
- (218) Hu, Q.; Wang, M.; Cho, M. S.; Wang, C.; Nick, A. M.; Thiagarajan, P.; Aung, F. M.; Han, X.; Sood, A. K.; Afshar-Kharghan, V. *BBA Clin* **2016**, *6*, 76–81.
- (219) Hishikawa, D.; Hashidate, T.; Shimizu, T.; Shindou, H. *J Lipid Res* **2014**, *55*, 799–807.
- (220) Li, P.; Gawaz, M.; Chatterjee, M.; Lammerhofer, M. *Anal Chem* **2021**, *93*, 4342–4350.
- (221) Harayama, T.; Riezman, H. *Nat Rev Mol Cell Biol* **2018**, *19*, 281–296.
- (222) Shaikh, S. R.; Kinnun, J. J.; Leng, X.; Williams, J. A.; Wassall, S. R. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2015**, *1848*, NMR Spectroscopy for Atomistic Views of Biomembranes and Cell Surfaces, 211–219.

- (223) Lu, Y.; Hong, S.; Tjonahen, E.; Serhan, C. N. *J Lipid Res* **2005**, *46*, 790–802.
- (224) Zou, Y. et al. *Nature* **2020**, *585*, 603–608.
- (225) Jiang, X.; Stockwell, B. R.; Conrad, M. *Nat Rev Mol Cell Biol* **2021**, DOI: 10.1038/s41580-020-00324-8.
- (226) Liu, A.; Wu, Q.; Guo, J.; Ares, I.; Rodriguez, J. L.; Martinez-Larranaga, M. R.; Yuan, Z.; Anadon, A.; Wang, X.; Martinez, M. A. *Pharmacol Ther* **2019**, *195*, 54–84.
- (227) Harm, T.; Frey, M.; Dittrich, K.; Goldschmied, A.; Rohlfing, A. K.; Fu, X.; Brun, A.; Castor, T.; Rath, D.; Muller, K.; Lammerhofer, M.; Gawaz, M. *Thromb Haemost* **2023**, DOI: 10.1055/s-0043-1764353.
- (228) Gawaz, M. *JACC: Cardiovascular Imaging* **2016**, *9*, 855–857.
- (229) Yamaja-Setty, B.; Berger, M.; Stuart, M. *Biochemical and Biophysical Research Communications* **1987**, *148*, 528–533.
- (230) Serhan, C. N.; Chiang, N.; Dalli, J.; Levy, B. D. *Cold Spring Harb Perspect Biol* **2014**, *7*, a016311.
- (231) Olzmann, J. A.; Carvalho, P. *Nat Rev Mol Cell Biol* **2019**, *20*, 137–155.
- (232) Harm, T.; Dittrich, K.; Brun, A.; Fu, X.; Frey, M.; Petersen Uribe, A.; Schwarz, F. J.; Rohlfing, A. K.; Castor, T.; Geisler, T.; Rath, D.; Lammerhofer, M.; Gawaz, M. P. *Clin Res Cardiol* **2023**, DOI: 10.1007/s00392-023-02260-x.
- (233) Xiao, H.; Siddiqui, R. A.; Al-Hassani, M. H.; Sliva, D.; Kovacs, R. J. *Platelets* **2001**, *12*, 163–170.
- (234) Shah, S. H.; Sun, J.-L.; Stevens, R. D.; Bain, J. R.; Muehlbauer, M. J.; Pieper, K. S.; Haynes, C.; Hauser, E. R.; Kraus, W. E.; Granger, C. B.; Newgard, C. B.; Califf, R. M.; Newby, L. K. *American Heart Journal* **2012**, *163*, 844–850.e1.
- (235) Roy, J.; Adili, R.; Kulmacz, R.; Holinstat, M.; Das, A. *Journal of Pharmacology and Experimental Therapeutics* **2016**, *359*, 134–141.
- (236) Smith, R. D.; Kelly, C. N. M.; Fielding, B. A.; Hauton, D.; Silva, K. D. R. R.; Nydahl, M. C.; Miller, G. J.; Williams, C. M. *British Journal of Nutrition* **2003**, *90*, 597–606.
- (237) Sato, T.; Nakao, K.; Hashizume, T.; Fujii, T. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1987**, *931*, 157–164.
- (238) Deguchi, H.; Banerjee, Y.; Trauger, S.; Siuzdak, G.; Kalisiak, E.; Fernandez, J. A.; Hoang, L.; Tran, M.; Yegneswaran, S.; Elias, D. J.; Griffin, J. H. *Blood* **2015**, *126*, 1595–1600.
- (239) Petersen-Urbe, A.; Kremser, M.; Rohlfing, A.-K.; Castor, T.; Kolb, K.; Dicenta, V.; Emschermann, F.; Li, B.; Borst, O.; Rath, D.; Müller, K. A. L.; Gawaz, M. P. *International Journal of Molecular Sciences* **2021**, *22*, DOI: 10.3390/ijms222011179.

- 
- (240) Cebo, M. et al. *Blood* **2022**, *139*, 1722–1742.
- (241) Weber, C. *Circulation Research* **2005**, *96*, 612–616.
- (242) Wang, C.; Chen, W.; Shen, J. *Frontiers in Pharmacology* **2018**, *9*, DOI: 10.3389/fphar.2018.00641.
- (243) Barupal, D. K.; Fan, S.; Wancewicz, B.; Cajka, T.; Sa, M.; Showalter, M. R.; Baillie, R.; Tenenbaum, J. D.; Louie, G.; Alzheimer's Disease Neuroimaging, I.; Alzheimer's Disease Metabolomics, C.; Kaddurah-Daouk, R.; Fiehn, O. *Sci Data* **2018**, *5*, 180263.
- (244) Bagheri, M.; Tiwari, H. K.; Murillo, A. L.; Al-Tobasei, R.; Arnett, D. K.; Kind, T.; Barupal, D. K.; Fan, S.; Fiehn, O.; O'Connell, J.; Montasser, M.; Aslibekyan, S.; Irvin, M. R. *Lipids Health Dis* **2020**, *19*, 153.
- (245) Alshehry, Z. H. et al. *Circulation* **2016**, *134*, 1637–1650.
- (246) Mundra, P. A. et al. *JCI Insight* **2018**, *3*, DOI: 10.1172/jci.insight.121326.

## A Appendix

### A.1 Data processing parameters in MS-DIAL

Table A.1: Parameters used to process SWATH lipidomics data in positive and negative ion mode of prestudy samples

| parameter                                                                                 | positive ion mode                                         | negative ion mode                   |
|-------------------------------------------------------------------------------------------|-----------------------------------------------------------|-------------------------------------|
| <b>Data collection parameters</b>                                                         |                                                           |                                     |
| Retention time begin                                                                      | 0.5                                                       | 0.5                                 |
| Retention time end                                                                        | 13                                                        | 13                                  |
| Mass range begin                                                                          | 50                                                        | 50                                  |
| Mass range end                                                                            | 1250                                                      | 1050                                |
| <b>Centroid parameters</b>                                                                |                                                           |                                     |
| MS1 tolerance                                                                             | 0.01                                                      | 0.01                                |
| MS2 tolerance                                                                             | 0.05                                                      | 0.05                                |
| Peak detection-based                                                                      | True                                                      | True                                |
| <b>Isotope recognition</b>                                                                |                                                           |                                     |
| Maximum charged number                                                                    | 2                                                         | 2                                   |
| <b>Data processing</b>                                                                    |                                                           |                                     |
| Number of threads                                                                         | 4                                                         | 4                                   |
| <b>Peak detection parameters</b>                                                          |                                                           |                                     |
| Smoothing method                                                                          | LinearWeightedMovingAverage                               | LinearWeightedMovingAverage         |
| Smoothing level                                                                           | 3                                                         | 3                                   |
| Minimum peak width                                                                        | 5                                                         | 5                                   |
| Minimum peak height                                                                       | 4000                                                      | 2000                                |
| <b>Peak spotting parameters</b>                                                           |                                                           |                                     |
| Mass slice width                                                                          | 0.1                                                       | 0.1                                 |
| <b>Deconvolution parameters</b>                                                           |                                                           |                                     |
| Peak consideration                                                                        | Both                                                      | Both                                |
| Sigma window value                                                                        | 0.5                                                       | 0.5                                 |
| Exclude after precursor                                                                   | True                                                      | True                                |
| <b>MSP file and MS/MS identification setting</b>                                          |                                                           |                                     |
| MSP file                                                                                  | MS-DIAL program ver. 2.94\MSDIAL-LipidDBs-VS36-FiehnO.lbm |                                     |
| Retention time tolerance                                                                  | 0.5                                                       | 0.5                                 |
| Accurate mass tolerance (MS1)                                                             | 0.01                                                      | 0.01                                |
| Accurate mass tolerance (MS2)                                                             | 0.05                                                      | 0.05                                |
| Identification score cut off                                                              | 80                                                        | 80                                  |
| <b>Text file and post identification (retention time and accurate mass based) setting</b> |                                                           |                                     |
| Retention time tolerance                                                                  | 0.1                                                       | 0.1                                 |
| Accurate mass tolerance                                                                   | 0.01                                                      | 0.01                                |
| Identification score cut off                                                              | 85                                                        | 85                                  |
| <b>Advanced setting for identification</b>                                                |                                                           |                                     |
| Relative abundance cut off                                                                | 0                                                         | 0                                   |
| Top candidate report                                                                      | True                                                      | True                                |
| <b>Adduct ion setting</b>                                                                 |                                                           |                                     |
|                                                                                           | [M+H] <sup>+</sup>                                        | [M-H] <sup>-</sup>                  |
|                                                                                           | [M+NH <sub>4</sub> ] <sup>+</sup>                         | [M-H <sub>2</sub> O-H] <sup>-</sup> |
|                                                                                           | [M+Na] <sup>+</sup>                                       | [M+FA-H] <sup>-</sup>               |
|                                                                                           | [M+H-H <sub>2</sub> O] <sup>+</sup>                       |                                     |

*continued*

| parameter                           | positive ion mode             | negative ion mode             |
|-------------------------------------|-------------------------------|-------------------------------|
| <b>Alignment parameters setting</b> |                               |                               |
| Reference file                      | pos mode\180625_PS-B1-101.abf | neg mode\180625_PS-B1-101.abf |
| Retention time tolerance            | 0.1                           | 0.1                           |
| MS1 tolerance                       | 0.025                         | 0.025                         |
| Retention time factor               | 0.5                           | 0.5                           |
| MS1 factor                          | 0.5                           | 0.5                           |
| Peak count filter                   | 2                             | 2                             |
| QC at least filter                  | False                         | False                         |
| <b>Tracking of isotope labels</b>   |                               |                               |
| Tracking of isotopic labels         | FALSE                         | FALSE                         |

Table A.2: Parameters used to process SWATH lipidomics data in positive and negative ion mode of main study samples

| parameter                         | positive ion mode           | negative ion mode           |
|-----------------------------------|-----------------------------|-----------------------------|
| <b>Data collection parameters</b> |                             |                             |
| Retention time begin              | 0.5                         | 0.5                         |
| Retention time end                | 13                          | 13                          |
| Mass range begin                  | 50                          | 50                          |
| Mass range end                    | 1250                        | 1050                        |
| MS2 mass range begin              | 5                           | 5                           |
| MS2 mass range end                | 1250                        | 1050                        |
| <b>Centroid parameters</b>        |                             |                             |
| MS1 tolerance                     | 0.01                        | 0.01                        |
| MS2 tolerance                     | 0.025                       | 0.025                       |
| <b>Isotope recognition</b>        |                             |                             |
| Maximum charged number            | 2                           | 2                           |
| <b>Data processing</b>            |                             |                             |
| Number of threads                 | 10                          | 10                          |
| <b>Peak detection parameters</b>  |                             |                             |
| Smoothing method                  | LinearWeightedMovingAverage | LinearWeightedMovingAverage |
| Smoothing level                   | 3                           | 3                           |
| Minimum peak width                | 5                           | 5                           |
| Minimum peak height               | 1000                        | 500                         |
| <b>Peak spotting parameters</b>   |                             |                             |
| Mass slice width                  | 0.1                         | 0.1                         |
| <b>Deconvolution parameters</b>   |                             |                             |
| Sigma window value                | 0.5                         | 0.5                         |
| MS2Dec amplitude cut off          | 0                           | 0                           |
| Exclude after precursor           | True                        | True                        |
| Keep isotope until                | 0.5                         | 0.5                         |
| Keep original precursor isotopes  | False                       | False                       |

*continued*

| parameter                                                                                 | positive ion mode                                 | negative ion mode                   |
|-------------------------------------------------------------------------------------------|---------------------------------------------------|-------------------------------------|
| <b>MSP file and MS/MS identification setting</b>                                          |                                                   |                                     |
| MSP file                                                                                  | MSDIAL ver.4.70 \Msp20210527163602_converted.lbm2 |                                     |
| Retention time tolerance                                                                  | 0.5                                               | 0.5                                 |
| Accurate mass tolerance (MS1)                                                             | 0.01                                              | 0.01                                |
| Accurate mass tolerance (MS2)                                                             | 0.05                                              | 0.05                                |
| Identification score cut off                                                              | 80                                                | 80                                  |
| Using retention time for scoring                                                          | False                                             | False                               |
| Using retention time for filtering                                                        | False                                             | False                               |
| <b>Text file and post identification (retention time and accurate mass based) setting</b> |                                                   |                                     |
| Retention time tolerance                                                                  | 0.15                                              | 0.15                                |
| Accurate mass tolerance                                                                   | 0.01                                              | 0.01                                |
| Identification score cut off                                                              | 85                                                | 85                                  |
| <b>Advanced setting for identification</b>                                                |                                                   |                                     |
| Relative abundance cut off                                                                | 0                                                 | 0                                   |
| Top candidate report                                                                      | False                                             | False                               |
| <b>Adduct ion setting</b>                                                                 |                                                   |                                     |
|                                                                                           | [M+H] <sup>+</sup>                                | [M-H] <sup>-</sup>                  |
|                                                                                           | [M+NH <sub>4</sub> ] <sup>+</sup>                 | [M-H <sub>2</sub> O-H] <sup>-</sup> |
|                                                                                           | [M+Na] <sup>+</sup>                               | [M+FA-H] <sup>-</sup>               |
|                                                                                           | [M+H-H <sub>2</sub> O] <sup>+</sup>               |                                     |
| <b>Alignment parameters setting</b>                                                       |                                                   |                                     |
| Reference file                                                                            | <i>injection40.abf</i>                            | <i>injection40.abf</i>              |
| Retention time tolerance                                                                  | 0.1                                               | 0.1                                 |
| MS1 tolerance                                                                             | 0.015                                             | 0.015                               |
| Retention time factor                                                                     | 0.5                                               | 0.5                                 |
| MS1 factor                                                                                | 0.5                                               | 0.5                                 |
| Peak count filter                                                                         | 5                                                 | 5                                   |
| N% detected in at least one group                                                         | 0                                                 | 0                                   |
| Remove feature based on peak height fold-change                                           | True                                              | True                                |
| Sample max / blank average                                                                | 5                                                 | 5                                   |
| Sample average / blank average                                                            | 5                                                 | 5                                   |
| Keep identified and annotated metabolites                                                 | True                                              | True                                |
| Keep removable features and assign the tag for checking                                   | True                                              | True                                |
| Gap filling by compulsion                                                                 | True                                              | True                                |
| <b>Tracking of isotope labels</b>                                                         |                                                   |                                     |
| Tracking of isotopic labels                                                               | FALSE                                             | FALSE                               |
| <b>Ion mobility</b>                                                                       |                                                   |                                     |
| Ion mobility data                                                                         | FALSE                                             | FALSE                               |



## A.2 Inter-batch feature alignment

Table A.3: Summary of linear equations applied in retention time adjustment of batches prior inter-batch feature alignment

| cohort     | batch | positive ion mode |             |                | negative ion mode |             |                |
|------------|-------|-------------------|-------------|----------------|-------------------|-------------|----------------|
|            |       | slope             | y-intercept | R <sup>2</sup> | slope             | y-intercept | R <sup>2</sup> |
| prestudy   | B01   | 1.0049            | -0.0021     | 1.0000         | 1.0096            | -0.0144     | 1.0000         |
|            | B02   | 0.9973            | -0.0026     | 1.0000         | 0.9944            | 0.0040      | 1.0000         |
|            | B03   | 0.9978            | 0.0048      | 1.0000         | 0.9961            | 0.0103      | 1.0000         |
| main study | B06   | 0.9915            | 0.0455      | 1.0000         | 0.9910            | 0.0432      | 1.0000         |
|            | B07   | 0.9942            | 0.0729      | 1.0000         | 0.9971            | 0.0610      | 0.9999         |
|            | B11   | 1.0034            | -0.0366     | 1.0000         | 1.0004            | -0.0390     | 0.9999         |
|            | B15   | 1.0145            | -0.1202     | 0.9999         | 1.0118            | -0.1052     | 0.9998         |
|            | B18   | 0.9760            | -0.0504     | 0.9998         | 0.9651            | -0.0080     | 0.9998         |
|            | B19   | 0.9898            | -0.1402     | 0.9998         | 0.9805            | -0.0967     | 0.9996         |
|            | B20   | 0.9859            | -0.1577     | 0.9998         | 0.9655            | -0.0847     | 0.9998         |
|            | B21   | 0.9870            | -0.1542     | 0.9998         | 0.9723            | -0.0942     | 0.9998         |

## A.3 Data curation of identified features

Table A.4: Lipid target addition to reference feature list after manual feature search and identification

| ion mode | lipid species (sum) | lipid species (w/ side chains) | formula       | neutral exact mass | adduct                 | precursor exact m/z | tR [min] | target list      |
|----------|---------------------|--------------------------------|---------------|--------------------|------------------------|---------------------|----------|------------------|
| pos      | CAR 14:0;O          |                                | C21H41NO5     | 387.2985           | [M+H] <sup>+</sup>     | 388.3058            | 0.82     | manually added   |
| pos      | CAR 14:2            |                                | C21H37NO4     | 367.2723           | [M+H] <sup>+</sup>     | 368.2796            | 0.77     | former "Unknown" |
| pos      | CAR 16:0;O          |                                | C23H45NO5     | 415.3298           | [M+H] <sup>+</sup>     | 416.3371            | 1.02     | manually added   |
| pos      | CAR 18:0;O          |                                | C25H49NO5     | 443.3611           | [M+H] <sup>+</sup>     | 444.3684            | 1.64     | manually added   |
| pos      | CAR 18:1;O          |                                | C25H47NO5     | 441.3454           | [M+H] <sup>+</sup>     | 442.3527            | 1.17     | manually added   |
| pos      | CAR 20:4            |                                | C27H45NO4     | 447.3349           | [M+H] <sup>+</sup>     | 448.3422            | 1.25     | manually added   |
| pos      | CAR 22:1            |                                | C29H55NO4     | 481.4131           | [M+H] <sup>+</sup>     | 482.4204            | 3.02     | manually added   |
| pos      | CAR 22:6            |                                | C29H45NO4     | 471.3349           | [M+H] <sup>+</sup>     | 472.3422            | 1.17     | manually added   |
| pos      | CAR 4:0             |                                | C11H21NO4     | 231.1471           | [M+H] <sup>+</sup>     | 232.1544            | 0.48     | manually added   |
| pos      | CAR 4:0;O           |                                | C11H21NO5     | 247.1420           | [M+H] <sup>+</sup>     | 248.1493            | 0.48     | manually added   |
| pos      | CAR 5:0             |                                | C12H23NO4     | 245.1627           | [M+H] <sup>+</sup>     | 246.1700            | 0.48     | manually added   |
| pos      | CAR 5:0;O           |                                | C12H23NO5     | 261.1576           | [M+H] <sup>+</sup>     | 262.1649            | 0.47     | manually added   |
| pos      | CAR 8:0             |                                | C15H29NO4     | 287.2097           | [M+H] <sup>+</sup>     | 288.2170            | 0.54     | manually added   |
| pos      | Cer 33:1;O2 d7;IS   | Cer 18:1;O2/15:0 (d7)          | C33H58D7NO3   | 530.5400           | [M+H-H2O] <sup>+</sup> | 513.5367            | 5.22     | manually added   |
| pos      | DG 16:0             | DG 8:0_8:0                     | C19H36O5      | 344.2563           | [M+NH4] <sup>+</sup>   | 362.2901            | 1.96     | manually added   |
| pos      | DG 16:0             | DG 8:0_8:0                     | C19H36O5      | 344.2563           | [M+Na] <sup>+</sup>    | 367.2455            | 1.96     | manually added   |
| pos      | DG 18:0             | DG 10:0_8:0                    | C21H40O5      | 372.2876           | [M+NH4] <sup>+</sup>   | 390.3214            | 2.71     | manually added   |
| pos      | DG 18:0             | DG 10:0_8:0                    | C21H40O5      | 372.2876           | [M+Na] <sup>+</sup>    | 395.2768            | 2.71     | manually added   |
| pos      | DG 20:0             | DG 10:0_10:0                   | C23H44O5      | 400.3189           | [M+NH4] <sup>+</sup>   | 418.3528            | 3.13     | manually added   |
| pos      | DG 20:0             | DG 10:0_10:0                   | C23H44O5      | 400.3189           | [M+Na] <sup>+</sup>    | 423.3082            | 3.13     | manually added   |
| pos      | DG 38:3             |                                | C41H74O5      | 646.5536           | [M+NH4] <sup>+</sup>   | 664.5875            | 6.41     | manually added   |
| pos      | DG 33:1 d7;IS       | DG 15:0_18:1(d7)               | C36H61D7O5    | 587.5506           | [M+Na] <sup>+</sup>    | 610.5399            | 6.53     | manually added   |
| neg      | FA 10:0;O           |                                | C10H20O3      | 188.1412           | [M-H] <sup>-</sup>     | 187.1340            | 0.68     | manually added   |
| neg      | FA 16:0;O           |                                | C16H32O3      | 272.2351           | [M-H] <sup>-</sup>     | 271.2279            | 2.22     | manually added   |
| neg      | FA 18:0;O           |                                | C18H36O3      | 300.2664           | [M-H] <sup>-</sup>     | 299.2592            | 2.91     | manually added   |
| neg      | FA 20:0;O           |                                | C20H40O3      | 328.2977           | [M-H] <sup>-</sup>     | 327.2905            | 3.23     | manually added   |
| neg      | FA 22:0;O           |                                | C22H44O3      | 356.3290           | [M-H] <sup>-</sup>     | 355.3218            | 4.08     | manually added   |
| neg      | FA 4:0;O            |                                | C4H8O3        | 104.0473           | [M-H] <sup>-</sup>     | 103.0401            | 0.47     | manually added   |
| neg      | FA 6:0;O            |                                | C6H12O3       | 132.0786           | [M-H] <sup>-</sup>     | 131.0714            | 0.48     | manually added   |
| neg      | FA 8:0;O            |                                | C8H16O3       | 160.1099           | [M-H] <sup>-</sup>     | 159.1027            | 0.53     | manually added   |
| neg      | FA 17:4;O           |                                | C17H26O3      | 278.1882           | [M-H] <sup>-</sup>     | 277.1810            | 0.94     | manually added   |
| pos      | LPC 18:0            | LPC 0:0/18:0                   | C26H54NO7P    | 523.3638           | [M+Na] <sup>+</sup>    | 546.3531            | 2.04     | manually added   |
| pos      | LPC 18:1            | LPC 0:0/18:1                   | C26H52NO7P    | 521.3481           | [M+Na] <sup>+</sup>    | 544.3374            | 1.45     | manually added   |
| pos      | LPC 18:2            | LPC 0:0/18:2                   | C26H50NO7P    | 519.3325           | [M+Na] <sup>+</sup>    | 542.3218            | 1.07     | manually added   |
| pos      | LPC 18:1 d7;IS      | LPC 0:0/18:1 (d7)              | C26H45D7NO7P  | 528.3921           | [M+Na] <sup>+</sup>    | 551.3814            | 1.44     | manually added   |
| pos      | LPC 18:1 d7;IS      | LPC 18:1/0:0 (d7)              | C26H45D7NO7P  | 528.3921           | [M+Na] <sup>+</sup>    | 551.3814            | 1.57     | manually added   |
| neg      | LPI 16:0            |                                | C25H49O12P    | 572.2962           | [M-H] <sup>-</sup>     | 571.2890            | 1.11     | manually added   |
| neg      | LPI 18:1            |                                | C27H51O12P    | 598.3118           | [M-H] <sup>-</sup>     | 597.3046            | 1.21     | manually added   |
| pos      | LSM 14:2;O2         |                                | C19H39N2O5P   | 406.2597           | [M+H] <sup>+</sup>     | 407.2670            | 0.76     | manually added   |
| pos      | LSM 18:1;O2         |                                | C23H49N2O5P   | 464.3379           | [M+H] <sup>+</sup>     | 465.3452            | 0.93     | manually added   |
| neg      | FA 18:2;O           |                                | C18H32O3      | 296.2351           | [M-H] <sup>-</sup>     | 295.2279            | 1.14     | manually added   |
| pos      | PC 18:0             | PC 10:0_8:0                    | C26H52NO8P    | 537.3431           | [M+H] <sup>+</sup>     | 538.3504            | 1.28     | manually added   |
| pos      | PC 20:0             | PC 10:0_10:0                   | C28H56NO8P    | 565.3744           | [M+H] <sup>+</sup>     | 566.3817            | 1.95     | manually added   |
| pos      | PC 24:0             | PC 16:0_8:0                    | C32H64NO8P    | 621.4370           | [M+H] <sup>+</sup>     | 622.4443            | 3.21     | manually added   |
| pos      | PC 30:4             | PC 10:0_20:4                   | C38H68NO8P    | 697.4683           | [M+H] <sup>+</sup>     | 698.4756            | 3.28     | manually added   |
| pos      | PC 33:1             |                                | C41H80NO8P    | 745.5622           | [M+Na] <sup>+</sup>    | 768.5515            | 5.17     | manually added   |
| pos      | PC 34:3             |                                | C42H78NO8P    | 755.5465           | [M+Na] <sup>+</sup>    | 778.5358            | 4.58     | manually added   |
| pos      | PC 36:3             |                                | C44H82NO8P    | 783.5778           | [M+Na] <sup>+</sup>    | 806.5671            | 5.31     | manually added   |
| pos      | PC 40:1             |                                | C48H94NO8P    | 843.6717           | [M+H] <sup>+</sup>     | 844.6790            | 7.50     | manually added   |
| pos      | PC 33:1 d7;IS       | PC 15:0_18:1(d7)               | C41H73D7NO8P  | 752.6061           | [M+Na] <sup>+</sup>    | 775.5954            | 5.15     | manually added   |
| neg      | PE 38:2             | PE 18:1_20:1                   | C43H82NO8P    | 771.5778           | [M-H] <sup>-</sup>     | 770.5706            | 6.35     | manually added   |
| neg      | PE 42:5             | PE 20:4_22:1                   | C47H84NO8P    | 821.5935           | [M-H] <sup>-</sup>     | 820.5863            | 6.36     | manually added   |
| neg      | PE 42:9             | PE 22:5_20:4                   | C47H76NO8P    | 813.5309           | [M-H] <sup>-</sup>     | 812.5237            | 4.95     | manually added   |
| neg      | PI 40:5             | PI 20:4_20:1                   | C49H85O13P    | 912.5728           | [M-H] <sup>-</sup>     | 911.5656            | 4.89     | manually added   |
| neg      | PS 34:1             |                                | C40H76NO10P   | 761.5207           | [M-H] <sup>-</sup>     | 760.5135            | 5.00     | former "Unknown" |
| neg      | PS 34:2             | PS 18:2_16:0                   | C40H74NO10P   | 759.5050           | [M-H] <sup>-</sup>     | 758.4978            | 4.45     | manually added   |
| neg      | PS 36:3             | PS 18:2_18:1                   | C42H76NO10P   | 785.5207           | [M-H] <sup>-</sup>     | 784.5135            | 4.58     | manually added   |
| neg      | PS 40:5             | PS 18:0_22:5                   | C46H80NO10P   | 837.5520           | [M-H] <sup>-</sup>     | 836.5448            | 5.19     | manually added   |
| neg      | PS 42:1             | PS 18:1_24:0                   | C48H92NO10P   | 873.6459           | [M-H] <sup>-</sup>     | 872.6387            | 7.45     | manually added   |
| pos      | PS 33:1 d7;IS       | PS 15:0_18:1(d7)               | C39H67D7NO10P | 754.5490           | [M+H] <sup>+</sup>     | 755.5563            | 4.60     | manually added   |

*continued*

| ion mode | lipid species (sum) | lipid species (w/ side chains)         | formula       | neutral exact mass | adduct               | precursor exact m/z | tR [min] | target list      |
|----------|---------------------|----------------------------------------|---------------|--------------------|----------------------|---------------------|----------|------------------|
| pos      | SM 36:2;O2 d9;IS    | SM 18:1;2O/18:1 (d9)                   | C41H72D9N2O6P | 737.6397           | [M+Na] <sup>+</sup>  | 760.6290            | 4.77     | manually added   |
| pos      | ST 27:1;O d7;IS     |                                        | C27H39OD7     | 393.3988           | [M+H] <sup>+</sup>   | 394.4061            | 4.75     | manually added   |
| neg      | FA 10:0             | TG 10:0_10:0_10:0<br>TG 10:0_10:0_10:0 | C10H20O2      | 172.1463           | [M-H] <sup>-</sup>   | 171.1391            | 1.07     | manually added   |
| neg      | FA 12:0             |                                        | C12H24O2      | 200.1776           | [M-H] <sup>-</sup>   | 199.1704            | 1.58     | manually added   |
| neg      | FA 22:0             |                                        | C22H44O2      | 340.3341           | [M-H] <sup>-</sup>   | 339.3269            | 4.89     | manually added   |
| neg      | FA 24:0             |                                        | C24H48O2      | 368.3654           | [M-H] <sup>-</sup>   | 367.3582            | 5.65     | manually added   |
| neg      | FA 26:0             |                                        | C26H52O2      | 396.3967           | [M-H] <sup>-</sup>   | 395.3895            | 6.44     | manually added   |
| pos      | TG 30:0             |                                        | C33H62O6      | 554.4546           | [M+Na] <sup>+</sup>  | 577.4439            | 5.62     | manually added   |
| pos      | TG 30:0             |                                        | C33H62O6      | 554.4546           | [M+NH4] <sup>+</sup> | 572.4885            | 5.64     | manually added   |
| pos      | TG 32:0             |                                        | C35H66O6      | 582.4859           | [M+NH4] <sup>+</sup> | 600.5198            | 6.35     | manually added   |
| neg      | FA 20:3;O4          |                                        | C20H34O6      | 370.2355           | [M-H] <sup>-</sup>   | 369.2283            | 3.06     | manually added   |
| pos      | Coenzyme Q10(ox)    |                                        | C59H90O4      | 862.6839           | [M+NH4] <sup>+</sup> | 880.7178            | 9.40     | manually added   |
| pos      | Coenzyme Q10(red)   |                                        | C59H90O4      | 864.6984           | [M+NH4] <sup>+</sup> | 882.7334            | 9.40     | former "Unknown" |
| neg      | FA 20:2             |                                        | C20H36O2      | 308.2715           | [M-H] <sup>-</sup>   | 307.2643            | 3.30     | manually added   |
| pos      | alpha-Tocopherol    |                                        | C29H50O2      | 430.3811           | [M+H] <sup>+</sup>   | 431.3884            | 5.17     | former "Unknown" |

## A.4 Targeted data extraction in MultiQuant

Table A.5: Summary of linear equations applied in retention time adjustment of batches prior targeted data extraction

| cohort     | batch | positive ion mode |             |                | negative ion mode |             |                |
|------------|-------|-------------------|-------------|----------------|-------------------|-------------|----------------|
|            |       | slope             | y-intercept | R <sup>2</sup> | slope             | y-intercept | R <sup>2</sup> |
| prestudy   | B01   | 1.0049            | -0.0021     | 1.0000         | 1.0096            | -0.0144     | 1.0000         |
|            | B02   | 0.9973            | -0.0026     | 1.0000         | 0.9944            | 0.0040      | 1.0000         |
|            | B03   | 0.9978            | 0.0048      | 1.0000         | 0.9961            | 0.0103      | 1.0000         |
| main study | B01   | 1.0245            | -0.0494     | 0.9999         | 1.0213            | 0.0725      | 0.9999         |
|            | B02   | 0.9880            | -0.0341     | 0.9998         | 1.0212            | 0.0690      | 0.9999         |
|            | B03   | 0.9879            | -0.0249     | 0.9998         | 1.0194            | 0.0751      | 0.9999         |
|            | B04   | 1.0660            | 0.1226      | 0.9995         | 1.0615            | 0.0742      | 0.9998         |
|            | B05   | 1.0065            | -0.1342     | 0.9995         | 0.9917            | -0.0844     | 0.9998         |
|            | B06   | 1.0085            | -0.0603     | 0.9999         | 1.0069            | -0.0412     | 0.9999         |
|            | B07   | 1.0054            | -0.0785     | 0.9999         | 1.0006            | -0.0546     | 0.9998         |
|            | B08   | 1.0140            | 0.0545      | 0.9996         | 1.0083            | 0.0436      | 0.9999         |
|            | B09   | 1.0111            | 0.0359      | 0.9996         | 1.0037            | 0.0363      | 0.9999         |
|            | B10   | 1.0068            | 0.0313      | 0.9995         | 1.0010            | 0.0343      | 0.9999         |
|            | B11   | 0.9971            | 0.0260      | 0.9999         | 0.9975            | 0.0385      | 0.9999         |
|            | B12   | 1.0216            | 0.0961      | 0.9997         | 1.0288            | 0.0760      | 0.9999         |
|            | B13   | 1.0221            | 0.0977      | 0.9997         | 1.0297            | 0.0774      | 0.9999         |
|            | B14   | 1.0261            | 0.1139      | 0.9997         | 1.0141            | 0.0658      | 0.9999         |
|            | B15   | 0.9869            | 0.1102      | 0.9999         | 0.9954            | 0.0753      | 0.9998         |
|            | B16   | 1.0248            | 0.1317      | 0.9997         | 1.0404            | 0.0698      | 0.9999         |
|            | B17   | 1.0224            | 0.1224      | 0.9997         | 1.0365            | 0.0749      | 0.9999         |
|            | B18   | 1.0259            | 0.0416      | 0.9999         | 1.0439            | -0.0269     | 0.9999         |
|            | B19   | 1.0145            | 0.1140      | 0.9999         | 1.0294            | 0.0555      | 0.9999         |
|            | B20   | 1.0187            | 0.1270      | 0.9999         | 1.0384            | 0.0646      | 0.9999         |
|            | B21   | 1.0173            | 0.1201      | 0.9999         | 1.0347            | 0.0606      | 0.9999         |
|            | B22   | 1.0400            | 0.0990      | 0.9996         | 1.0473            | 0.0520      | 0.9998         |

Table A.6: Parameters used for targeted data extraction in MultiQuant

| parameter                   | <i>continued</i>          |        |
|-----------------------------|---------------------------|--------|
|                             | value                     | unit   |
| Gaussian Smooth             | 1.5                       | points |
| Expected RT                 | <i>feature-specific</i>   | min    |
| RT half window              | 6.0                       | sec    |
| Report largest peak         | <i>unchecked checkbox</i> | -      |
| Min. peak width             | 5                         | points |
| Min. peak height            | 50                        | points |
| Noise percentage            | 50                        | %      |
| Baseline subtraction window | 2.00                      | min    |
| Peak splitting              | 1                         | points |

## A.5 Lipidome alterations in dependence on disease severity

|                                                 | All (n = 139)  | CCS Troponin –<br>(n = 100, 71.9%) | ACS Troponin +<br>(n = 39, 28.1%) | P-value |
|-------------------------------------------------|----------------|------------------------------------|-----------------------------------|---------|
| Male, n (%)                                     | 101 (72.7%)    | 71 (71.0%)                         | 30 (76.9%)                        | 0.477   |
| Age, years (mean ± SD)                          | 70.6 (±11.2)   | 70.4 (±10.4)                       | 71.2 (±13.1)                      | 0.310   |
| Body mass index (mean ± SD)                     | 27.1 (±4.7)    | 27.3 (±4.7)                        | 26.3 (±4.8)                       | 0.170   |
| Cardiovascular risk factors                     |                |                                    |                                   |         |
| Arterial hypertension, n (%)                    | 109 (78.4%)    | 80 (80.0%)                         | 29 (74.4%)                        | 0.473   |
| Hyperlipidaemia, n (%)                          | 64 (46.0%)     | 52 (52.0%)                         | 12 (30.8%)                        | 0.023   |
| Diabetes mellitus, n (%)                        | 41 (29.5%)     | 27 (27.0%)                         | 14 (35.9%)                        | 0.307   |
| Current smoking, n (%)                          | 23 (16.5%)     | 17 (17.0%)                         | 6 (15.4%)                         | 0.817   |
| Ex smoking >6 months, n (%)                     | 23 (16.5%)     | 17 (17.0%)                         | 6 (15.4%)                         | 0.817   |
| Obesity, n (%)                                  | 65 (46.8%)     | 54 (54.0%)                         | 11 (28.2%)                        | 0.006   |
| Atrial fibrillation, n (%)                      | 33 (23.7%)     | 24 (24.0%)                         | 9 (23.1%)                         | 0.908   |
| Previous CABG, n (%)                            | 13 (9.4%)      | 11 (11.0%)                         | 2 (5.1%)                          | 0.260   |
| Previous MI, n (%)                              | 29 (20.9%)     | 25 (25.0%)                         | 4 (10.3%)                         | 0.043   |
| Renal function (GFR) (mean ± SD)                | 74.9 (± 30.63) | 73.0 (± 25.7)                      | 79.6 (± 40.6)                     | 0.577   |
| Medication on admission                         |                |                                    |                                   |         |
| Acetylsalicylic acid, n (%)                     | 112 (80.6%)    | 83 (83.0%)                         | 29 (74.4%)                        | 0.257   |
| Clopidogrel, n (%)                              | 36 (25.9%)     | 34 (34.0%)                         | 2 (5.1%)                          | 0.001   |
| Ticagrelor, n (%)                               | 44 (31.7%)     | 31 (31.0%)                         | 12 (30.8%)                        | 0.505   |
| Prasugrel, n (%)                                | 9 (6.5%)       | 6 (6.0%)                           | 3 (7.7%)                          | 0.720   |
| Cangrelor, n (%)                                | 1 (0.7%)       | 0 (0.0%)                           | 1 (2.6%)                          | 0.110   |
| Oral anticoagulants, n (%)                      | 33 (23.7%)     | 28 (28.0%)                         | 5 (12.8%)                         | 0.048   |
| Angiotensin-converting enzyme inhibitors, n (%) | 71 (51.1%)     | 53 (53.0%)                         | 18 (46.2%)                        | 0.468   |
| Angiotensin II receptor antagonists, n (%)      | 32 (23.0%)     | 25 (25.0%)                         | 7 (17.9%)                         | 0.366   |
| Aldosterone antagonists, n (%)                  | 28 (20.1%)     | 24 (24.0%)                         | 4 (10.3%)                         | 0.056   |
| Ca channel antagonists, n (%)                   | 38 (27.3%)     | 31 (31.0%)                         | 7 (17.9%)                         | 0.111   |
| β-blockers, n (%)                               | 89 (64.0%)     | 67 (67.0%)                         | 22 (56.4%)                        | 0.246   |
| Diuretics                                       | 54 (38.8%)     | 42 (42.0%)                         | 12 (30.8%)                        | 0.218   |
| Statins, n (%)                                  | 92 (66.2%)     | 76 (76.0%)                         | 16 (41.0%)                        | 0.001   |
| Lipid profile parameters                        |                |                                    |                                   |         |
| LDL-cholesterol (mg/dL) (mean ± SD)             | 96.0 (±32.9)   | 94.1 (±33.7)                       | 101.2 (±30.5)                     | 0.193   |
| HDL-cholesterol (mg/dL) (mean ± SD)             | 47.6 (±16.1)   | 48.8 (±16.7)                       | 44.4 (±14.0)                      | 0.225   |
| Triglycerides (mg/dL) (mean ± SD)               | 150.4 (±90.2)  | 153.4 (±89.6)                      | 142.3 (±92.5)                     | 0.351   |
| Total cholesterol (mg/mL) (mean ± SD)           | 161.2 (±43.6)  | 161.0 (±42.7)                      | 161.7 (±46.7)                     | 0.936   |
| Platelets (10 <sup>9</sup> /L) (mean ± SD)      | 225.1 (±84.8)  | 229.9 (±72.7)                      | 230.8 (±110.5)                    | 0.717   |
| Disease                                         |                |                                    |                                   |         |
| Stable angina                                   | 75 (54.0%)     | 75 (75.0%)                         | –                                 |         |
| Unstable angina                                 | 25 (18.0%)     | 25 (25.0%)                         | –                                 |         |
| NSTEMI                                          | 29 (20.9%)     | –                                  | 29 (74.4%)                        |         |
| STEMI                                           | 10 (7.2%)      | –                                  | 10 (25.6%)                        |         |
| Follow-up events (12 months)                    |                |                                    |                                   |         |
| Death                                           | 7 (5.0%)       | 6 (6.0%)                           | 1 (2.6%)                          | 0.375   |
| Myocardial infarction                           | 0 (0.0%)       | 0 (0.0%)                           | 0 (0.0%)                          |         |
| Stroke                                          | 0 (0.0%)       | 0 (0.0%)                           | 0 (0.0%)                          |         |

Figure A.1: Summary of the baseline characteristics of the prestudy patients cohort focusing on disease severity. Student's t-test was applied to normally distributed data while non-normally distributed data were analyzed by the Mann-Whitney-U-test. Pearson Chi-square was used for discrete data calculations. Mean values and odds ratio are given with 95% confidence interval. Reprinted with permission from [214] (Copyright 2021, European Society of Cardiology).

## A.6 Lipidome alterations in dependence on statin treatment

|                                                 | All            | Statin treatment | Statin naïve    | p-Value      |
|-------------------------------------------------|----------------|------------------|-----------------|--------------|
|                                                 | (n = 105)      | (n = 71; 67.6%)  | (n = 34; 32.4%) |              |
| Male, n (%)                                     | 77 (73)        | 54 (51.4)        | 23 (21.9)       | 0.362        |
| Age, y (mean ± SD)                              | 71.1 (± 10.7)  | 71.7 (± 10.6)    | 69.8 (± 11.1)   | 0.386        |
| Body mass index (mean ± SD)                     | 27 (± 4.4)     | 27.3 (± 4.3)     | 26.6 (± 4.7)    | 0.538        |
| Cardiovascular risk factors                     |                |                  |                 |              |
| Arterial hypertension, n (%)                    | 85 (81)        | 59 (83.1)        | 26 (76.5)       | 0.418        |
| Hyperlipidemia, n (%)                           | 50 (47.6)      | 39 (54.9)        | 11 (32.3)       | <b>0.030</b> |
| Diabetes mellitus, n (%)                        | 29 (27.6)      | 24 (33.8)        | 5 (14.7)        | <b>0.041</b> |
| Current smoking, n (%)                          | 19 (18.1)      | 13 (18.3)        | 6 (17.6)        | 0.934        |
| Ex smoking > 6 mo, n (%)                        | 14 (13.3)      | 9 (12.6)         | 5 (14.7)        | 0.775        |
| Obesity, n (%)                                  | 53 (50.5)      | 39 (54.9)        | 14 (41.2)       | 0.187        |
| Atrial fibrillation, n (%)                      | 24 (22.9)      | 16 (22.5)        | 8 (23.5)        | 0.910        |
| Previous CABG, n (%)                            | 11 (10.5)      | 10 (14.1)        | 1 (2.9)         | 0.081        |
| Previous MI, n (%)                              | 21 (20)        | 19 (26.8)        | 2 (5.9)         | <b>0.012</b> |
| Renal function (GFR) (mean ± SD)                | 72.4 (± 29.4)  | 71.3 (± 23.9)    | 74.7 (± 38.9)   | 0.643        |
| Medication on admission                         |                |                  |                 |              |
| Statins, n (%)                                  | 71 (67.6)      |                  |                 |              |
| Acetylsalicylic acid, n (%)                     | 88 (83.8)      | 62 (87.3)        | 26 (76.5)       | 0.158        |
| Clopidogrel, n (%)                              | 30 (28.6)      | 22 (31)          | 8 (23.5)        | 0.429        |
| Ticagrelor, n (%)                               | 36 (34.3)      | 28 (39.4)        | 8 (23.5)        | 0.108        |
| Prasugrel, n (%)                                | 6 (5.7)        | 3 (4.2)          | 3 (8.8)         | 0.342        |
| Cangrelor, n (%)                                | 1 (1)          | 0                | 1 (2.9)         | 0.147        |
| Oral anticoagulants, n (%)                      | 27 (25.7)      | 22 (31)          | 5 (14.7)        | <b>0.074</b> |
| Angiotensin-converting enzyme inhibitors, n (%) | 55 (52.4)      | 40 (56.3)        | 15 (44.1)       | 0.241        |
| Angiotensin II receptor antagonists, n (%)      | 25 (23.8)      | 20 (28.2)        | 5 (14.7)        | 0.130        |
| Aldosterone antagonists, n (%)                  | 21 (20)        | 16 (22.5)        | 5 (14.7)        | 0.348        |
| Ca channel antagonists, n (%)                   | 30 (28.6)      | 23 (32.4)        | 7 (20.5)        | 0.210        |
| β-blockers, n (%)                               | 72 (68.6)      | 53 (74.6)        | 19 (55.9)       | 0.053        |
| Diuretics, n (%)                                | 41 (39)        | 35 (49.3)        | 6 (17.6)        | <b>0.002</b> |
| Lipid profile parameters                        |                |                  |                 |              |
| LDL-cholesterol (mg/dL) (mean ± SD)             | 96.5 (± 34.7)  | 89.4 (± 31.4)    | 111.2 (± 34.6)  | <b>0.001</b> |
| HDL-cholesterol (mg/dL) (mean ± SD)             | 48.6 (± 17.6)  | 46.6 (± 15.8)    | 52.3 (± 18.4)   | 0.235        |
| Triglycerides (mg/dL) (mean ± SD)               | 145.2 (± 85.1) | 148.1 (± 76.3)   | 142.8 (± 102.7) | 0.879        |
| Total cholesterol (mg/mL) (mean ± SD)           | 161.3 (± 45.9) | 151.8 (± 43.1)   | 183.4 (± 38.3)  | <b>0.002</b> |
| Platelets (mean ± SD)                           | 235.8 (± 75.8) | 213.4 (± 74.1)   | 254.6 (± 83.1)  | 0.104        |
| Disease                                         |                |                  |                 |              |
| Chronic coronary syndrome, n (%)                | 60 (57.1)      | 43 (60.6)        | 17 (50)         | <b>0.007</b> |
| Unstable angina, n (%)                          | 19 (18.1)      | 16 (22.5)        | 3 (8.8)         | 0.088        |
| NSTEMI, n (%)                                   | 22 (21)        | 12 (16.9)        | 10 (29.4)       | 0.141        |
| STEMI, n (%)                                    | 4 (3.8)        | 0 (0)            | 4 (11.8)        | <b>0.003</b> |
| Follow-up events                                |                |                  |                 |              |
| Death, n (%)                                    | 6 (5.7)        | 4 (5.6)          | 2 (5.9)         | 0.959        |
| Myocardial Infarction, n (%)                    | 0              | 0                | 0               |              |
| Stroke, n (%)                                   | 0              | 0                | 0               |              |

Abbreviations: CABG, coronary artery bypass graft; HDL, high-density lipoprotein; LDL, low-density lipoprotein; NSTEMI, non-ST-elevation myocardial infarction; SD, standard deviation; STEMI, ST-elevation myocardial infarction.

Note: Normally distributed data were analyzed using Student's t-test, nonnormally distributed data were computed using the Mann-Whitney U-test. Mean values are presented as mean, lower and upper 95% confidence intervals. Discrete data were calculated using Chi-square test (Pearson). All significant p-values ( $p < 0.05$ ) of comparison between treatment groups were highlighted bold.

Figure A.2: Baseline characteristics of the prestudy patients cohort focusing on statin treatment. Reprinted with permission from [227] (Copyright 2023, Thieme).

## A.7 Cardiovascular risk assessment in CAD

|                                                  | All<br>(n=1057)      | No adverse events<br>(n=985; 93.2%) | Composite Endpoint<br>(n=72; 6.8%) | p-value      |
|--------------------------------------------------|----------------------|-------------------------------------|------------------------------------|--------------|
| Female, n (%)                                    | 318 (30.1)           | 303 (30.8)                          | 15 (20.8)                          | 0.067        |
| Age, years (mean $\pm$ SD)                       | 70.2 ( $\pm$ 11.6)   | 69.9 ( $\pm$ 11.5)                  | 74.2 ( $\pm$ 12)                   | <b>0.004</b> |
| Body mass index (mean $\pm$ SD)                  | 27.5 ( $\pm$ 5.1)    | 27.6 ( $\pm$ 5.1)                   | 27.0 ( $\pm$ 4.6)                  | 0.366        |
| <b>Cardiovascular risk factors</b>               |                      |                                     |                                    |              |
| Arterial hypertension, n (%)                     | 945 (89.4)           | 879 (89.1)                          | 66 (91.7)                          | 0.518        |
| Hyperlipidemia, n (%)                            | 815 (77.1)           | 757 (76.9)                          | 58 (80.6)                          | 0.470        |
| Diabetes mellitus, n (%)                         | 319 (30.2)           | 294 (29.9)                          | 25 (34.7)                          | 0.384        |
| Current smoking, n (%)                           | 197 (18.6)           | 187 (18.9)                          | 10 (13.9)                          | 0.280        |
| Ex Smoking > 6 mo, n (%)                         | 228 (21.6)           | 208 (21.1)                          | 20 (27.8)                          | 0.188        |
| Obesity, n (%)                                   | 257 (24.3)           | 238 (24.2)                          | 19 (26.4)                          | 0.671        |
| Atrial Fibrillation, n (%)                       | 246 (23.3)           | 223 (22.6)                          | 23 (31.9)                          | 0.073        |
| Previous CABG, n (%)                             | 47 (4.5)             | 40 (4.1)                            | 4 (9.7)                            | <b>0.025</b> |
| Previous MI, n (%)                               | 226 (21.4)           | 207 (21)                            | 19 (26.4)                          | 0.283        |
| Renal function (GFR) (mean $\pm$ SD)             | 80.1 ( $\pm$ 40.7)   | 80.4 ( $\pm$ 41.4)                  | 75.9 ( $\pm$ 29.5)                 | 0.232        |
| <b>Medication on admission</b>                   |                      |                                     |                                    |              |
| Statins, n (%)                                   | 855 (80.1)           | 795 (80.7)                          | 60 (83.3)                          | 0.521        |
| Acetylsalicylic acid, n (%)                      | 931 (88.1)           | 871 (88.4)                          | 60 (83.3)                          | 0.198        |
| Clopidogrel, n (%)                               | 460 (43.5)           | 432 (43.9)                          | 28 (38.9)                          | 0.429        |
| Ticagrelor, n (%)                                | 223 (21.1)           | 216 (21.9)                          | 17 (23.6)                          | 0.726        |
| Prasugrel, n (%)                                 | 157 (14.9)           | 149 (15.1)                          | 8 (11.1)                           | 0.359        |
| Cangrelor, n (%)                                 | 1 (0.1)              | 1 (0.1)                             | 0 (0)                              | 0.787        |
| Oral anticoagulants, n (%)                       | 232 (22)             | 210 (21.3)                          | 22 (30.6)                          | 0.068        |
| Angiotensin-converting enzyme inhibitors, n (%)  | 373 (35.3)           | 349 (35.4)                          | 24 (33.3)                          | 0.740        |
| Angiotensin II receptor antagonists, n (%)       | 498 (47.1)           | 465 (47.2)                          | 33 (45.8)                          | 0.850        |
| Angiotensin receptor-neprilysin inhibitor, n (%) | 12 (1.4)             | 11 (1.1)                            | 1 (1.4)                            | 0.718        |
| Aldosterone antagonists, n (%)                   | 239 (22.6)           | 219 (22.5)                          | 20 (27.8)                          | 0.267        |
| SGLT-2 inhibitors, n (%)                         | 60 (5.7)             | 56 (5.7)                            | 4 (5.6)                            | 0.765        |
| Ca channel antagonists, n (%)                    | 369 (34.9)           | 344 (34.9)                          | 25 (34.7)                          | 0.995        |
| $\beta$ -blockers, n (%)                         | 720 (68.1)           | 663 (67.3)                          | 57 (79.2)                          | 0.053        |
| Diuretics, n (%)                                 | 377 (35.7)           | 347 (35.2)                          | 30 (41.7)                          | 0.259        |
| <b>Lipid profile parameters</b>                  |                      |                                     |                                    |              |
| LDL-cholesterol (mg/dL) (mean $\pm$ SD)          | 106 ( $\pm$ 44.3)    | 108.1 ( $\pm$ 60.4)                 | 105.9 ( $\pm$ 43.1)                | 0.775        |
| HDL-cholesterol (mg/dL) (mean $\pm$ SD)          | 46.8 ( $\pm$ 19.8)   | 46.8 ( $\pm$ 20)                    | 45.6 ( $\pm$ 15.5)                 | 0.541        |
| Triglycerides (mg/dL) (mean $\pm$ SD)            | 145.7 ( $\pm$ 108.3) | 147.3 ( $\pm$ 110.5)                | 121.1 ( $\pm$ 63.6)                | <b>0.004</b> |
| Total cholesterol (mg/mL) (mean $\pm$ SD)        | 165.7 ( $\pm$ 46.4)  | 165.7 ( $\pm$ 44.9)                 | 165.2 ( $\pm$ 65.2)                | 0.954        |
| <b>Platelets (mean <math>\pm</math> SD)</b>      | 226.5 ( $\pm$ 67.6)  | 226.5 ( $\pm$ 66.5)                 | 225.8 ( $\pm$ 81)                  | 0.944        |
| <b>Disease</b>                                   |                      |                                     |                                    |              |
| Chronic coronary syndrome, n (%)                 | 396 (37.5)           | 371 (37.7)                          | 25 (24.5)                          | 0.619        |
| Unstable Angina, n (%)                           | 192 (18.2)           | 185 (18.8)                          | 7 (9.7)                            | 0.054        |
| NSTEMI, n (%)                                    | 326 (30.8)           | 295 (30)                            | 31 (43.1)                          | <b>0.020</b> |
| STEMI, n (%)                                     | 93 (8.8)             | 87 (8.8)                            | 6 (8.3)                            | 0.884        |
| Peripheral artery disease, n (%)                 | 50 (4.7)             | 47 (4.8)                            | 3 (4.2)                            | 0.815        |

Figure A.3: Baseline characteristics of the main study large-scale CAD cohort [Harm et al. 2023, submitted manuscript].

| 72 Months Follow-Up          | Ischemic Endpoint (n=51) | Bleeding Endpoint (n=21) |
|------------------------------|--------------------------|--------------------------|
| Cardiac Death, n (%)         | 17 (33.3)                |                          |
| Myocardial Infarction, n (%) | 21 (41.2)                |                          |
| Stroke, n (%)                | 13 (25.5)                |                          |
| Major Bleeding, n (%)        |                          | 21 (100)                 |

Figure A.4: Summary of follow-up events in the main study large-scale CAD cohort [Harm et al. 2023, submitted manuscript].



## List of Figures

|      |                                                                                                                                                     |    |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1  | Plasma membrane and cellular structure of platelets . . . . .                                                                                       | 2  |
| 1.2  | Resting and activated platelet . . . . .                                                                                                            | 4  |
| 1.3  | Role of platelets in haemostasis . . . . .                                                                                                          | 5  |
| 1.4  | Role of platelets in thrombosis . . . . .                                                                                                           | 7  |
| 1.5  | Platelet induced progression of atherosclerotic plaques . . . . .                                                                                   | 8  |
| 1.6  | Lipid categories based on LIPID MAPS . . . . .                                                                                                      | 10 |
| 1.7  | Structure of PC(16:0/18:1(9 <i>Z</i> )) and SM(d18:1(4 <i>E</i> )(1OH,3OH)18:1(9 <i>Z</i> )) . . .                                                  | 11 |
| 1.8  | Lipid biosynthesis . . . . .                                                                                                                        | 13 |
| 1.9  | Scheme of ion formation by ESI . . . . .                                                                                                            | 18 |
| 1.10 | Scheme of the Sciex TripleTOF 5600 . . . . .                                                                                                        | 20 |
| 1.11 | Structure of a quadrupole mass analyzer . . . . .                                                                                                   | 21 |
| 1.12 | Mathieu stability diagram . . . . .                                                                                                                 | 21 |
| 1.13 | Illustration of targeted data acquisition approaches . . . . .                                                                                      | 24 |
| 1.14 | Design of DIA experiments using SWATH . . . . .                                                                                                     | 27 |
| 1.15 | Exemplified QC chart . . . . .                                                                                                                      | 31 |
| 3.1  | Batch sequence design . . . . .                                                                                                                     | 44 |
| 3.2  | Inter-batch feature alignment procedure . . . . .                                                                                                   | 48 |
| 4.1  | Prestudy quality control charts . . . . .                                                                                                           | 59 |
| 4.2  | Batch-specific coefficient of variation in main study batches . . . . .                                                                             | 60 |
| 4.3  | Main study quality control charts . . . . .                                                                                                         | 61 |
| 4.4  | Flowchart of the finally developed data processing workflow . . . . .                                                                               | 63 |
| 4.5  | Linear retention time correlation . . . . .                                                                                                         | 65 |
| 4.6  | Starting Excel sheet with instructions for the application of the VBA alignment tool . . . . .                                                      | 66 |
| 4.7  | Pop-up windows to define alignment parameters . . . . .                                                                                             | 66 |
| 4.8  | Results of inter-batch feature alignment from three prestudy batches . . . . .                                                                      | 68 |
| 4.9  | Distribution of zeta values over all aligned identified features . . . . .                                                                          | 69 |
| 4.10 | Misalignment of PE 36:5 isomers . . . . .                                                                                                           | 70 |
| 4.11 | Alignment parameter evaluation . . . . .                                                                                                            | 72 |
| 4.12 | Beneficial effect of inter-batch feature alignment on lipid identification . . . . .                                                                | 74 |
| 4.13 | Development of proportion and number of identified lipids among all detected features in main study data by inter-batch feature alignment . . . . . | 76 |
| 4.14 | Homologic series of LPC species in positive ion mode . . . . .                                                                                      | 80 |
| 4.15 | Manual curation of automated lipid identification: Alignment spotting map and XIC . . . . .                                                         | 81 |
| 4.16 | Manual curation of automated lipid identification: Compound search panel . .                                                                        | 82 |
| 4.17 | Identification curation resulted in annotation adjustment by reducing identification level . . . . .                                                | 83 |

|                                                                                                                                                                                                     |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.18 Data curation of unknown features: Misalignments . . . . .                                                                                                                                     | 87  |
| 4.19 Data curation of unknown features: Misaligned unknown feature . . . . .                                                                                                                        | 88  |
| 4.20 Data curation of unknown features: Output of feature fidelity evaluation . . . . .                                                                                                             | 89  |
| 4.21 Data curation of unknown features: Result of feature fidelity evaluation . . . . .                                                                                                             | 89  |
| 4.22 Prestudy reference target list . . . . .                                                                                                                                                       | 90  |
| 4.23 Lipid categories and lipid classes of identified lipids of the prestudy reference list . . . . .                                                                                               | 91  |
| 4.24 Result summary of feature list alignment from prestudy and main study datasets . . . . .                                                                                                       | 92  |
| 4.25 Decrease of features in reference list through data curation . . . . .                                                                                                                         | 93  |
| 4.26 Lipid categories and lipid classes of identified lipids of the main study reference list . . . . .                                                                                             | 94  |
| 4.27 Retention time adjustment from reference to batch-specific values . . . . .                                                                                                                    | 95  |
| 4.28 Performance of the retention time adjustment shown as accuracy and precision of all features . . . . .                                                                                         | 97  |
| 4.29 Proportion of missing values among all samples and features in prestudy and main study data . . . . .                                                                                          | 98  |
| 4.30 Overview of metrics of variation for raw and normalized data in box-whisker plots in prestudy data . . . . .                                                                                   | 101 |
| 4.31 Reduction of inter-batch CV by data normalization compared to raw data determined in QC samples of prestudy data . . . . .                                                                     | 102 |
| 4.32 Overview of metrics of variation for raw and normalized data in box-whisker plots in main study data . . . . .                                                                                 | 103 |
| 4.33 PLS-DA plots of raw and CCMN-, PQN-, and SERRF-normalized data . . . . .                                                                                                                       | 105 |
| 4.34 Reduction of inter-batch CV by data normalization compared to raw data determined in QC samples of main study data . . . . .                                                                   | 106 |
| 4.35 Volcano plot of identified lipids . . . . .                                                                                                                                                    | 111 |
| 4.36 Radar chart of significantly altered lipids . . . . .                                                                                                                                          | 112 |
| 4.37 Boxplots of significantly upregulated lipids in ACS compared to CCS and healthy controls . . . . .                                                                                             | 114 |
| 4.38 Pathway hypothesis of platelet lipidome alterations involving MCFA-PCs . . . . .                                                                                                               | 116 |
| 4.39 Volcano plot illustrating lipidome alterations depending on statin treatment . . . . .                                                                                                         | 118 |
| 4.40 Heat maps of TG and PC lipid classes . . . . .                                                                                                                                                 | 119 |
| 4.41 Score plot of OPLS-DA analysis of lipidome alterations in dependence on statin-treatment in ACS and CCS subcohorts . . . . .                                                                   | 120 |
| 4.42 PCA of the identified platelet lipidome from the CAD cohort indicating separation of patients due to ischemic, bleeding, or no cardiovascular events . . . . .                                 | 122 |
| 4.43 Volcano plots differentiating eventless CAD patients from patients suffering adverse ischemic or bleeding events . . . . .                                                                     | 123 |
| 4.44 Evaluation of platelet lipidome alterations in dependence on APT in the clinical prestudy cohort and effects of <i>ex vivo</i> ACKR3/CXCR7 agonist treatment of CAD platelet samples . . . . . | 128 |

|     |                                                                                                                                                          |     |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.1 | Postulated metabolic pathway alterations in chronic and acute coronary syndrome and the effect of statin medication on the platelet lipid metabolism . . | 132 |
| A.1 | Summary of the baseline characteristics of the prestudy patients cohort focusing on disease severity . . . . .                                           | 153 |
| A.2 | Baseline characteristics of the prestudy patients cohort focusing on statin treatment . . . . .                                                          | 154 |
| A.3 | Baseline characteristics of the main study large-scale CAD cohort . . . . .                                                                              | 155 |
| A.4 | Summary of follow-up events in the main study large-scale CAD cohort . . .                                                                               | 156 |

## List of Tables

|     |                                                                                                                                                                                                                                                                                                                                                                                                                 |    |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Levels of lipid notation with increasing degree of structural information . . .                                                                                                                                                                                                                                                                                                                                 | 11 |
| 1.2 | Performance characteristics of typically used mass analyzers . . . . .                                                                                                                                                                                                                                                                                                                                          | 19 |
| 1.3 | Tools for processing of untargeted LC-MS/MS lipidomics data providing lipid identification and quantification adopted from LIPID MAPS homepage based on Ni et al. [161] . . . . .                                                                                                                                                                                                                               | 33 |
| 3.1 | Concentration of calibrants and internal standards in stock solution and six-point calibration series . . . . .                                                                                                                                                                                                                                                                                                 | 42 |
| 3.2 | MS parameters . . . . .                                                                                                                                                                                                                                                                                                                                                                                         | 43 |
| 3.3 | Design of MS (TOF) and MS/MS (SWATH) experiments . . . . .                                                                                                                                                                                                                                                                                                                                                      | 44 |
| 3.4 | List of SOPs established for the large-scale study . . . . .                                                                                                                                                                                                                                                                                                                                                    | 46 |
| 3.5 | Steps of lipidomics data processing performed in pre- and main study data sets with three and 22 acquisition batches, respectively . . . . .                                                                                                                                                                                                                                                                    | 52 |
| 4.1 | Summary of validation results showing the linear range, calibration parameters, as well as LOD and LOQ (quantified with respective IS). LOD and LOQ were determined in neat solution as $S/N \approx 3$ and $S/N \approx 10$ , respectively, multiplied by the calibrant specific process efficiency. Further results were obtained under consideration of endogenous amounts of the calibrant lipids . . . . . | 54 |
| 4.2 | Summary of validation results for within-batch and between-batch (i.e. three different days) precision and accuracy in QC samples spiked with three concentration levels of calibrants. Results were obtained under consideration of endogenous amounts of the calibrant lipids . . . . .                                                                                                                       | 55 |
| 4.3 | Summary of validation results for recovery (RE), matrix effect (ME), and process efficiency (PE) for calibrants calculated as the mean $\pm$ standard deviation for three concentration levels . . . . .                                                                                                                                                                                                        | 57 |
| 4.4 | Coefficient of variation of internal standards determined from raw peak intensities in QC samples regularly injected over the sequences of prestudy batches; mean CV with standard deviation over all three batches is also given . . . . .                                                                                                                                                                     | 58 |
| 4.5 | Confusion matrix showing true-positive and true-negative as well as false-positive and false-negative alignments among all features of Batches 01, 02, and 03 . . . . .                                                                                                                                                                                                                                         | 71 |
| 4.6 | Parameter combinations to evaluate the effect of parameter levels on alignment result . . . . .                                                                                                                                                                                                                                                                                                                 | 72 |
| 4.7 | Summary of reference feature lists after MS-DIAL data processing and inter-batch feature alignment in both positive and negative ion mode for prestudy and main study data . . . . .                                                                                                                                                                                                                            | 78 |
| 4.8 | Data curation of identified lipids in the prestudy dataset presented individually for positive and negative ion mode . . . . .                                                                                                                                                                                                                                                                                  | 84 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                              |     |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.9  | Data curation of identified lipids in the main study dataset combined for positive and negative ion mode . . . . .                                                                                                                                                                                                                                                                                           | 85  |
| 4.10 | Summary of PLS-DA results for raw and normalized data . . . . .                                                                                                                                                                                                                                                                                                                                              | 104 |
| 4.11 | Overview of all features in the final datamatrices of prestudy and main study cohorts . . . . .                                                                                                                                                                                                                                                                                                              | 107 |
| 4.12 | Assignment of IS to lipid classes for quantification by one-point calibration with lipid class specific IS . . . . .                                                                                                                                                                                                                                                                                         | 108 |
| 4.13 | Lipid concentrations of significantly up- and downregulated lipids in ACS patients compared to CCS. Concentrations were determined by relative quantification to lipid class specific internal standards and are given as median values with interquartile range (IQR) among each group in pmol/10 <sup>6</sup> cells. Reprinted with permission from [214] (Copyright 2021, European Society of Cardiology) | 113 |
| 4.14 | Summary of significantly upregulated lipids comparing patients with ischemic (n=51) or bleeding events (n=21) with patients without events (n=985). Results were obtained by student's t-test with BH FDR correction ( $p < 0.05$ , $FDR \leq 5\%$ ) based on lipid concentrations of the identified lipidome [232]                                                                                          | 124 |
| 4.15 | Hazard ratios (HR) from Cox proportional hazard models indicating the independent association of significantly upregulated lipids with the cardiovascular risk of ischemic or bleeding events . . . . .                                                                                                                                                                                                      | 125 |
| A.1  | Parameters used to process SWATH lipidomics data in positive and negative ion mode of prestudy samples . . . . .                                                                                                                                                                                                                                                                                             | 146 |
| A.2  | Parameters used to process SWATH lipidomics data in positive and negative ion mode of main study samples . . . . .                                                                                                                                                                                                                                                                                           | 147 |
| A.3  | Summary of linear equations applied in retention time adjustment of batches prior inter-batch feature alignment . . . . .                                                                                                                                                                                                                                                                                    | 149 |
| A.4  | Lipid target addition to reference feature list after manual feature search and identification . . . . .                                                                                                                                                                                                                                                                                                     | 150 |
| A.5  | Summary of linear equations applied in retention time adjustment of batches prior targeted data extraction . . . . .                                                                                                                                                                                                                                                                                         | 152 |
| A.6  | Parameters used for targeted data extraction in MultiQuant . . . . .                                                                                                                                                                                                                                                                                                                                         | 152 |