UCLA
UCLA Electronic Theses and Dissertations

Title

Highly Multiplexed Mass Spectrometry Based assay for Lipid-related Proteins Quantification

Permalink

https://escholarship.org/uc/item/01w2b2bg

Author

Kassem, Ahmad Mohammad

Publication Date

2022

Supplemental Material

https://escholarship.org/uc/item/01w2b2bg#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Highly Multiplexed Mass Spectrometry Based Assay for Lipid-related Proteins Quantification

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Biochemistry, Molecular and Structural Biology

by

Ahmad Mohammad Kassem

2023

© Copyright by

Ahmad Mohammad Kassem

2023

ABSTRACT OF THE THESIS

Highly Multiplexed Mass Spectrometry Based Assay for Lipid-related Proteins Quantification

by

Ahmad Mohammad Kassem

Master of Science in Biochemistry, Molecular and Structural Biology

University of California, Los Angeles, 2023

Professor Joseph A. Loo, Chair

The recent development of mass spectrometry-based lipidomics and chromatographic separation techniques has opened our eyes to the importance of many lipid species in various biochemical processes.

Lipidomics is now seen as an extension of metabolomics and a complementary field to genomics and proteomics. Understanding the interplay between all the fields of systems biology could be a compelling approach to examining disease-induced cellular changes and building comprehensive hypotheses for further investigation. In this study, we aim to develop a comprehensive mass spectrometry PRM targeting assay to measure the quantity of all human proteins known to play a role in lipid metabolism. We envision that this assay could add an extra layer of complementarity between proteomics and lipidomics. The highly multiplexed assay will allow us to explain shifts in cellular lipid profile by measuring the change in lipid-metabolizing proteins level. Examining both profiles will aid in hypothesis generation and open the opportunity for further biochemical investigation.

The thesis of Ahmad Mohammad Kassem is approved.

Jose A. Rodriguez

James A. Wohlschlegel

Joseph A. Loo, Committee Chair

University of California, Los Angeles

2023

Table of Contents

List of Figures

Introduction

The use of proteomics mass spectrometry has become an integral part of everyday biological and biochemical research. Mass spectrometry's unequivocal capability to identify and quantify proteins in a complex mixture has made it the ideal tool for large-scale discoveries. In a discovery shotgun proteomics experiment, an MS instrument operates in data-dependent acquisition (DDA) mode, and peptides are quantified by spectral counting [1] or integration of precursor ions' chromatogram peaks [2].

Alternatively, other isotopic-labeling methods such as ICAT [3] and SILAC [4] and chemical-labeling methods such as iTRAQ [5] have been introduced to enhance protein quantification. However, Mass spectrometry is not inherently quantitative because of thevariance in signal-dependence relative signal intensity of a peptide; hence, the stochastic approach of DDA results in limited quantification reproducibility for spectral counting and labeling approaches [1]. Additionally, DDA's bias to identify relatively high abundance peptides makes quantification reproducibility even worse for low abundant proteins.

Implementing selected/parallel reaction monitoring (SRM/PRM) in targeted proteomics has allowed for better quantification of a set of targeted proteins in a follow-up hypothesis-driven experiment [6]. SRM/PRM is characterized by high selectivity, cost and sample effectiveness, and large capacity for multiplexing. During an SRM/PRM experiment, a preselected group of peptides (based on a previous discovery experiment) is monitored to yield accurate, consistent, and reproducible quantification. Since its development, targeted proteomics has become the "golden"

method for clinicalbiomarker development [7,8,9].

In 2012, the SRM method was selected as the method of the year by the journal NatureMethods [10]. SRM takes advantage of triple quadrupole instruments to isolate and measure targeted transitions. Improvement of modern triple quadrupole instruments in scan speed and sensitivity allowed for highly multiplex analysis [11,12]. For example, Burgess and colleagues have used time-scheduled SRM to monitor 800 peptides with atotal of 2400 transitions using a long C18-LC gradient (~4hr) [11]. Despite that, SRM requires long and laborious method development time (The need to select for every transition) and has limited multiplexing capacity in large-scale studies.

SRM-like targeted acquisition such as PRM was pioneered on a high resolution, accurate mass platform based on quadrupole — orbitrap conjugate systems. In PRM, precursor ions are isolated in Q1, fragmented in Q2, and the full mass spectrum is recorded by high-resolution orbitrap [13,14,15]. The Acquisition of all transitions of the peptide and the high-speed cycle makes PRM an easier method to set up (no need to specify the parameters of every transition) and allows for a larger multiplexing capacitycompared to SRM.

Lipids are one of the four essential biomolecules in cells. There are thousands of lipid species that belong to eight main categories: fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, saccharolipids, polyketides, sterol lipids, and prenollipids [16]. Due to their highly complex chemical nature, lipids have always been very challenging to study and have historically received less attention than DNA and protein studies [17]. The scarcity of tools to reliable measure and study lipids has

created a gap in the biochemical research field. Not until recently were many lipid species such as ceramides considered inert species [18]. The recent development of mass spectrometry-based lipidomics and chromatographic separation techniques has openedour eyes to the importance of many lipid species in various biochemical processes [16].Lipidomics is now seen as an extension of metabolomics and a complementary field to genomics and proteomics; together, all these fields constitute systems biology.

Understanding the interplay between all the fields of systems biology could be a compelling approach to examining disease-induced cellular changes and building comprehensive hypotheses for further investigation. In this study, we aim to develop a comprehensive mass spectrometry PRM targeting assay to measure the quantity of allhuman proteins known to play a role in lipid metabolism. We envision that this assay could add an extra layer of complementarity between proteomics and lipidomics. The highly multiplexed assay will allow us to explain shifts in cellular lipid profile by measuring the change in lipid-metabolizing proteins level. Examining both profiles will aid in hypothesis generation and open the opportunity for further biochemical investigation.

Building the lipid-related protein list

To build the PRM-targeting assay, we must determine all proteins we should aim to target—deciding which proteins should be targeted proved tricky as many proteins could play an indirect role in lipid metabolism. In contrast, the role of other proteins in lipid metabolism is still debated. Therefore, we decided only to include proteins that playa direct role in lipid metabolism via direct lipid species degradation or

biosynthesis. Our choices for targeted proteins and their lipid-related pathways were based on Kegg's pathway database [19,20,21]. In total, we aimed to target 185 proteins; these proteins were categorized into nine different pathways (fig.1): (1) Fatty acids biosynthesis, (2) Fatty acids degradation, (3) Steroid biosynthesis, (4) Primary bile acid biosynthesis, (5) Steroid hormone biosynthesis, (6) Glycolipid metabolism, (7) Glycerophospholipid metabolism, (8) Sphingolipid metabolism, (9) Prostaglandin biosynthesis.

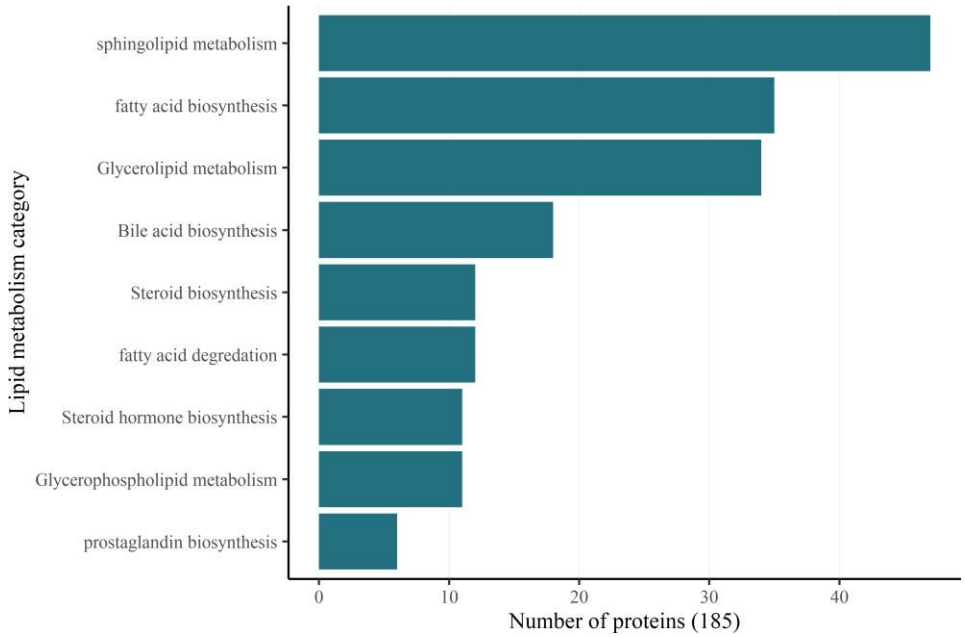

Fig1. Pathway distribution of targeted protein list. 185 proteins that play a role in lipid metabolism were selected for PRM targeting. All proteins were categorized by the lipid-related pathway theyare involved in.

To further investigate the properties of selected proteins, we applied gene ontology (GO) analysis. GO analysis showed that lipid metabolism is the most common biologicalprocess associated with these enzymes (fig.2a). For example, sphingolipid biosynthesisprocess, long fatty acid acyl chain biosynthesis, ceramide biosynthesis, and sterol metabolic processes. Additionally, molecular function analysis shows iron binding as themost common function in our protein list (fig.2b), hence, suggesting a potential role for an interplay between iron and lipid metabolism.

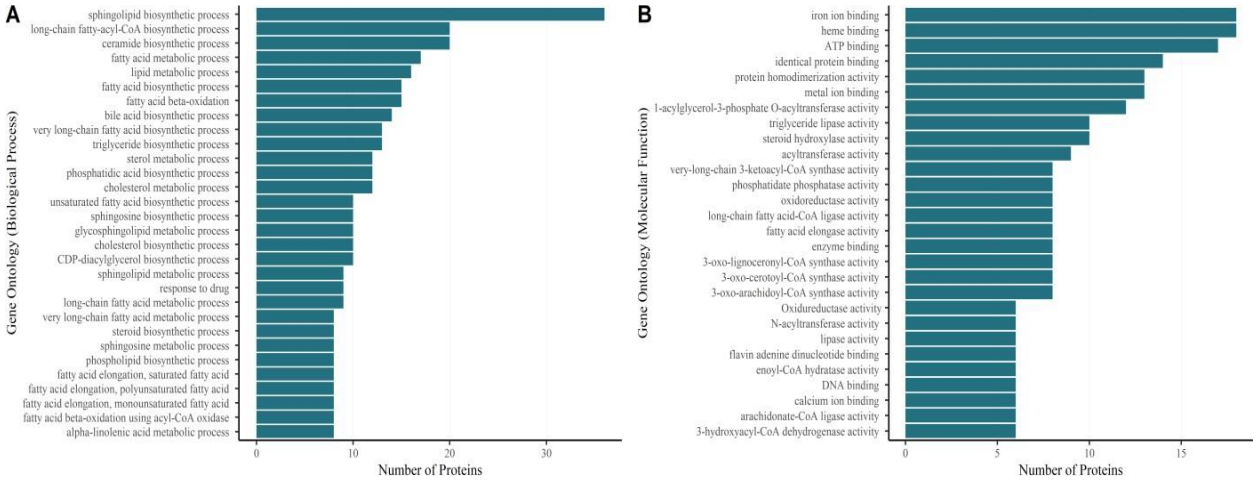

Fig2. Gene ontology analysis of biological processes and molecular functions of targetedproteins. (A) Gene ontology analysis of biological processes affirms the role of targeted proteins in lipid metabolism pathways. (B) Gene ontology analysis of targeted proteins' molecular function suggests an interplay between iron and lipid metabolism.

We also performed a GO analysis of protein localization (fig.3). The results show subcellular localization of most of our targeted proteins on membrane-bound organelles,including the endoplasmic reticulum, Golgi apparatus, mitochondria, lysosomes, peroxisomes, and cellular membrane. The localization of these proteins provides a convenient way for sample enrichment by subcellular fractionation techniques.

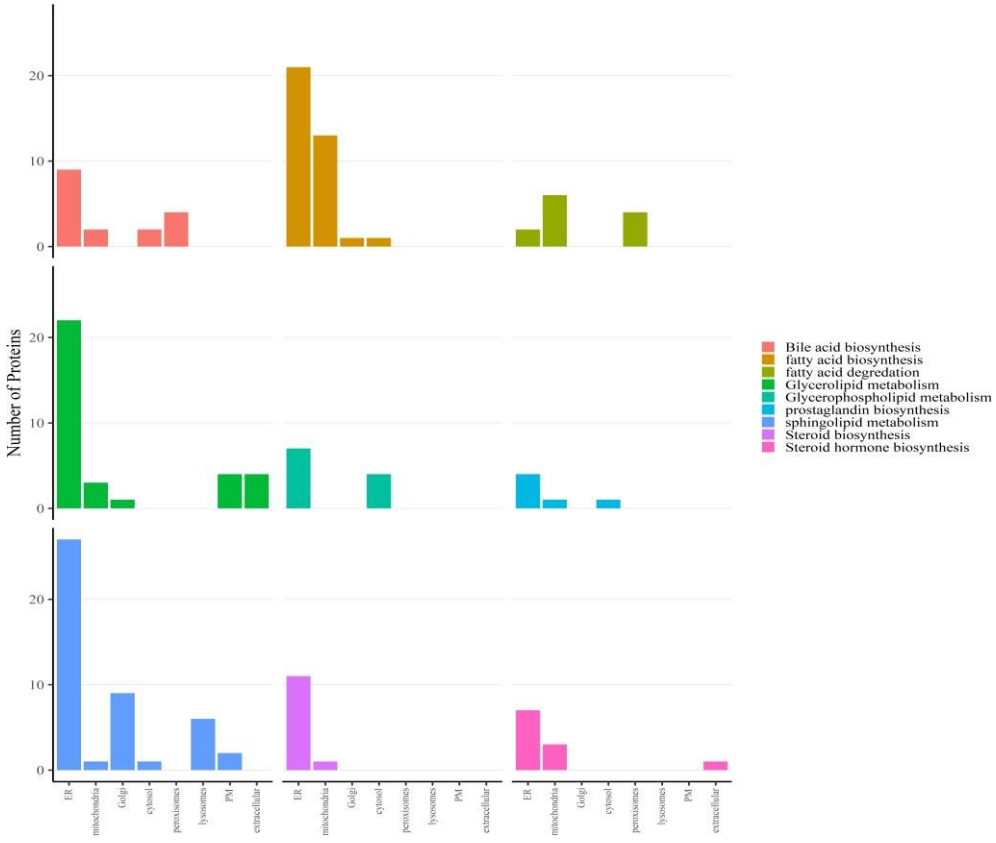

Fig 3. Subcellular localization of targeted proteins based on pathway. Localization analysisshows high localization of all targeted proteins on membrane bound organelles such as the ER, Golgi apparatus, mitochondria, peroxisomes and lysosomes regardless of pathway. Subcellular localization will be used for sample enrichment.

Picking the ideal biological system

One main problem in building our targeted assay was finding the cell line expressing all185 proteins at high rates for DDA detection. To determine the cell lines that express the targeted proteins, we searched the RNA expression library of the HUMAN Atlas Database [22]. The database contains the mRNA expression profile of 70 different commercially available cell lines. After the search, we decided to pick Hep3B (liver- derived cell line), U251-MG (glioblastoma-derived cell line), and OE19 (cardia/esophageal gastric junction-derived cell line). These cell lines were picked to maximize the number of identifiable proteins; they express 132 proteins. Trying to coverall 185 proteins using biological models would require >22 different cell lines (many of these proteins are tissue specific). Therefore, other methods such as in-vitro protein expression will be more time and cost-effective for identification purposes.

Sample Preparation and Mass spectrometry analysis

U251-MG and Hep3B cell lines were maintained at 37C, 5% CO2 in DMEM containing 4.5g glucose/L (GIBCO by Life Technologies) supplemented with 10% fetal bovine serum (Gemini Bioproducts), 1X antibiotic-antimycotic (GIBCO by Life Technologies) and 2mM L-glutamine (GIBCO by Life Technologies). OE19 cell line was maintained at 37C, 5% CO2 in RPMI1640 (GIBCO by Life Technologies) supplemented with 10% fetalbovine serum1X antibiotic-antimycotic and 2mM L-glutamine.

Six samples of each cell line (18 samples total) were harvested and fractionated based on Horton and Hordon's crude subcellular fractionation protocol, except for using 15

ug/ml digitonin for Hep3B fractionation. Membrane-bound organelle fractions of Trichloroacetic acid (Sigma-Aldrich no. T0699) were then added to membrane-bound fractionation samples to a final concentration of 20% and incubated on ice for 1 hr. Aftercentrifugation at 14,000 x g for 25 min at 4°C, pellets were washed with 500 µl of ice- cold acetone. After another centrifugation at 14,000 x g for 25 min at 4°C, the pellets were air-dried in a fume hood for 30 min. Protein pellets were re-dissolved in 8uM urea solution. Protein samples were reduced and alkylated. Using ethanol-activated SP3 magnetic beads, samples were digested in 0.1ug Lys-C and 0.4ug trypsin overnight.

The sample was analyzed by nanoflow liquid chromatography on an EASY-nLC 1000 system (Thermo) coupled online to a Fusion Lumos mass spectrometer (Thermo). Peptides were separated on a C18-reversed phase column (15 cm long, 75 µm inner diameter, packed in-house with ReproSil-Pur C18-AQ 1.9 µm resin (Dr. MaischGmbH)).We used a 140 min gradient from 2% to 80% acetonitrile in 0.5% formic acid at a flow of200 ml/min. The Thermo Fusion Lumos was operated with a Top10 MS/MS data- dependent acquisition method using HCD fragmentation.

MS data analysis

The raw files were processed with MaxQuant [23] (version 2.1.1.0) using the uniport- retrieved Human proteome database (UP000005640). Carbamidomethylation was setas a fixed modification, while methionine oxidation and protein N-acetylation were considered as variable modifications. The search was performed with an initial mass tolerance of 7 ppm mass accuracy for the precursor ion and 20 ppm for the MS/MS spectra in the HCD fragmentation mode.

Search results were filtered with a false discovery rate of 0.01 at the peptide and protein level.

The raw files were also processed by Fragpipe pipeline (version 18.0) using the uniport-retrieved Human proteome database (UP000005640) at default settings [24,25]. MS Fragger mass calibration and parameter optimization were applied with a precursor andproduct tolerance of 20 ppm for the MS/MS spectra. Carbamidomethylation was set as fixed modification, while methionine oxidation and protein N-acetylation were considered as variable modifications. Validation was performed by MSBooster, a deep learning algorithm based on predicted RT, predicted spectra, and correlated features.

Protein inference was performed by protein prophet. Search results were filtered with afalse discovery rate of 0.01 at the peptide and protein level. The msms and pepXML files were used to build a spectrum library using Skyline software [26]. The 185 targeted proteins were checked against both libraries (Fragpipe and MaxQuant libraries). Across all three cell lines, the fragpipe approach has identified1962 peptides correlating to 120 proteins compared to 1468 peptides correlating to 105 proteins (fig.4). The results show the exceeding capabilities of Fragpipe as a searching algorithm compared to MaxQuant. It is important to note that we only expected a maximum of ~132 protein identifications based on our previous Human Atlas Database search. A comparison of the contribution of each cell line to the overall number of identified proteins shows that a total of 824 peptides correlating to 76 proteins were identified in all three cell lines. Whereas 218, 162 and 45 peptides correlating to 7, 10 and 5 proteinswere exclusively identified in OE19, Hep3B and

U251-MG, respectively (fig.6A and fig.6B). The 34% increase in OE19-exculsively identified peptides compared to Hep3B does not correlate to higher number of identified proteins. This could be a result of higher expression or specific lipid-related proteins in OE19.

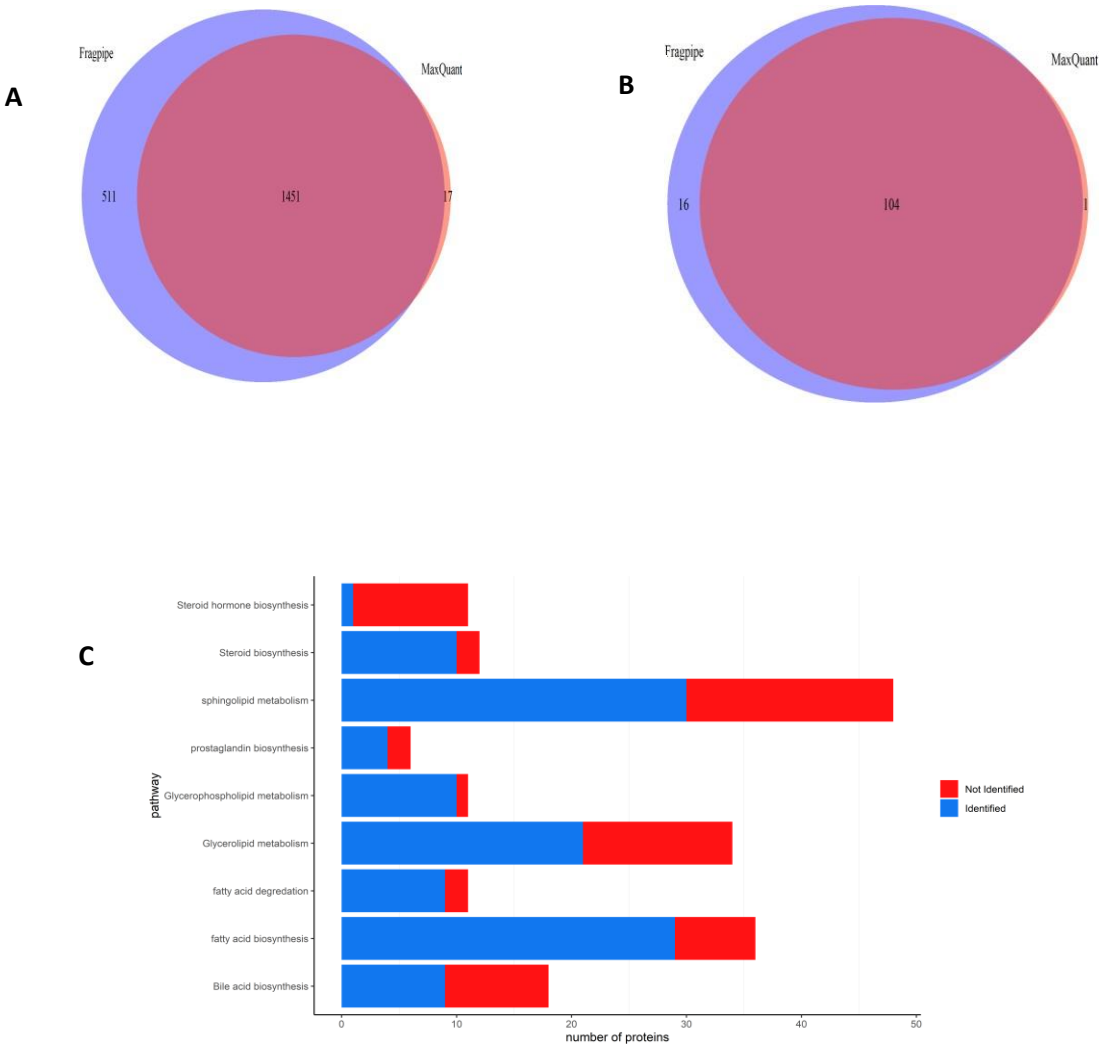

Fig4. Comparison of peptide and protein identification of Fragpipe and MaxQuant database searching engines. (A) Fragpipe and MaxQuant identified a total of 1979 peptides across all three cell lines(B) Fragpipe and MaxQuant identified 121 proteins from 132. (c) Overall view of number of identified vs. not identified proteins from every lipid-related pathway.

Skyline-assisted PRM-targeted method development

The first step of targeted method development was to select the candidate peptides that could be used for targeting and protein quantifications. The quantitypic peptides were selected using skyline targeted method development workflow. The peptides were filtered based on the following characteristics: (1) Peptide went through complete proteolysis (no missed cleavages) (2) Peptide is minimally prone to chemical modification (3) Peptide is minimally prone to biological modification (4) Peptide shows high ionization tendency (high number of PSMs) (5) Peptide is unique (6) Peptide has consistent ionization pattern (7) Peptide shows consistent C18-LC fractionation pattern. The peptide filtration step has decreased the number of targetable peptides across all three cell lines from 2019 initially identified peptides to 397, which correlates to 109 proteins out of the 120 identified proteins. The number of candidate peptides differs from one protein to another based on the size and abundance of the protein. Some proteins have only one targetable peptide, while others have up to nine targetable peptides (fig.5). Ideally, quantification at the protein level requires the quantification of atleast two separate and unique peptides; hence, proteins with one targetable peptide have "sub-optimal" quantification.

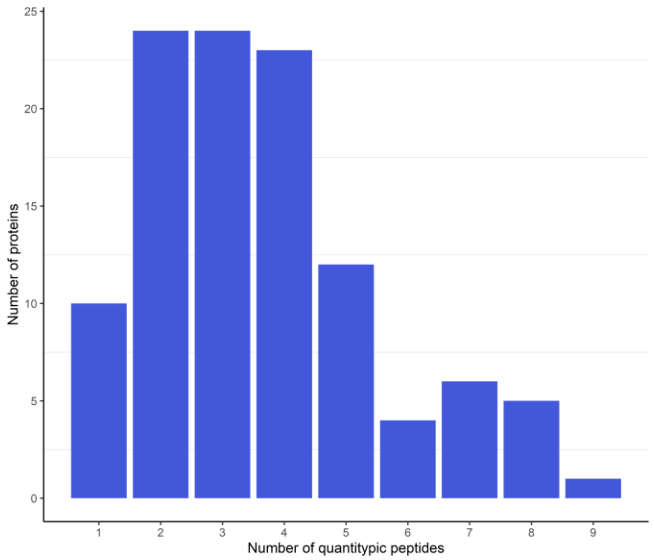

Fig5. Comparison of the number of Quantitypic peptides for all quantifiable proteins. Thenumber of quantitypic peptides for each protein differs between proteins based on the abundanceand size of protein. A higher number of quantitypic peptide correlates for more accurate and precise quantification of protein. It is “optimal” to measure at least two quantitypic per peptide foraccurate quantification.

Comparison of the contribution of cell lines in the identification of quantifiable proteins shows that 238 qantitypic peptides correlating to 65 quantifiable proteins were identified in all three cell lines. In contrast, 34, 17, and 11 quantitypic peptides correlating to 6,5, and 3 quantifiable proteins were exclusively identified in OE19, Hep3B and U251-MG,respectively (fig.6C and fig.6D).

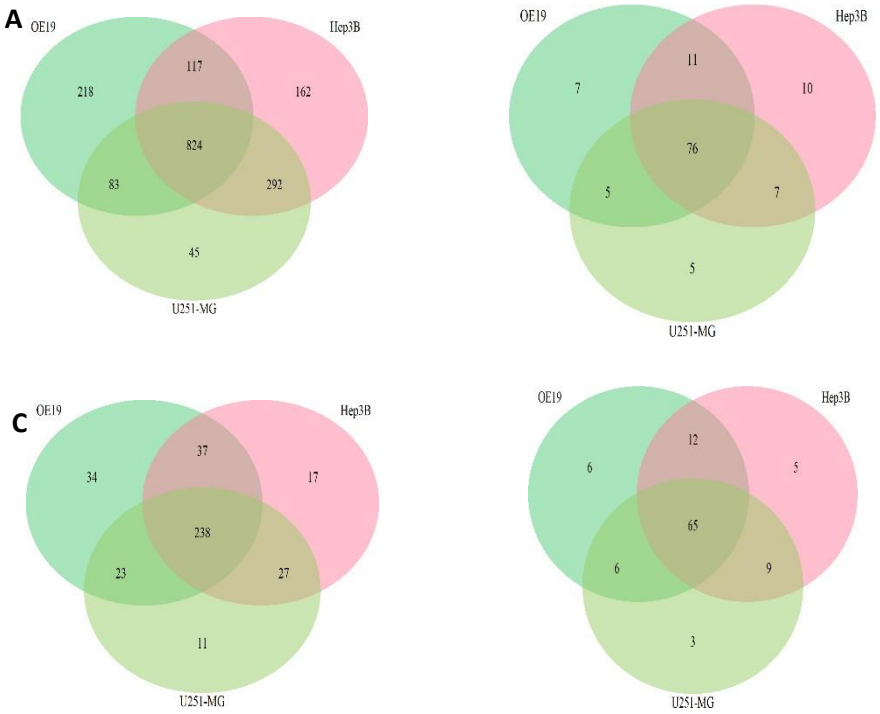

Fig6. Comparison of OE19, Hep3B and U251-MG contribution in peptide identification and quantitypic peptides selection. (A) Distribution of identified peptides among the three selected cell lines (B) Distribution of Identified proteins (C) Distribution of quantitypic peptidesacross the three cell lines. (D) Distribution of quantifiable proteins.

The second step of method development is assigning retention time for targeted peptides. The retention time of each peptide was measured as the average of all measured RTs in all 18 samples. The averaged retention time was converted to indexedretention time (iRT) to allow for data reproducibility across instruments and labs. The standard peptides for iRT calibration were peptides inherent to the samples. These peptides show high identification and high chromatographic reproducibility. Additionally, the iRT values based on the iRT-C18 standard curve (The most used iRT

curve) are known. Therefore, all averaged RTs were converted to iRT values based on the iRT- C18 calibration.

Finally, the method was built with a time scheduling of a 5-minutes window for each peptide in a 140min run. Time scheduling allows for the measurement of < 30 peptidesconcurrently (fig.7).

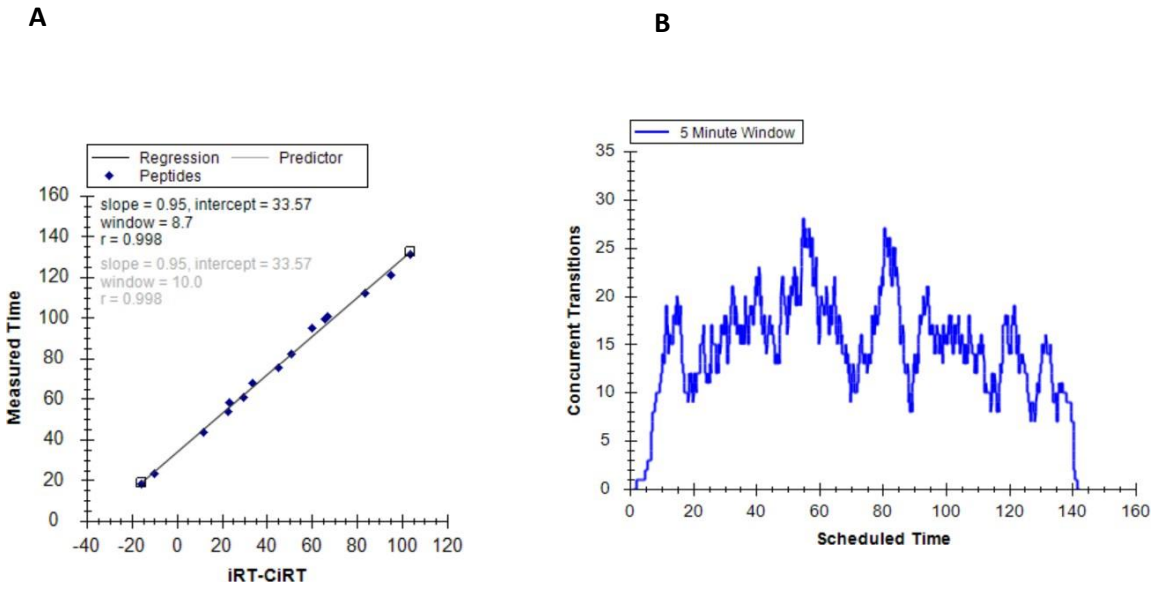

Fig7. iRT-C18 calibration curve and time scheduling for PRM-targeted method.

(A) 13 known peptides with high RT accuracy and precision with known iRT-C18 values were used to set a calibration curve and convert measured averaged retention times of tarted peptides to iRT.

(B) Time scheduling of PRM method allow for higher multiplexing capacity for protein quantification. Developed method ensure < 30 peptides to be simultaneously measured at anytime in a 5-min window of 140 minutes run

The Fusion Lumos instrument has a max scanning speed of 20Hz at 15000 resolutions,so you can monitor 20 peptides per second. If the chromatographic peak

width is 30 seconds and we need to monitor the peptide 10 times for accurate quantification, the total cycle time should be 2 seconds (30/10). So 20Hz * 3 = 60 peptides. Hence, a fusion Lumos instrument could theoretically measure 60 concurrent peptides and shouldn't struggle with our supplied method.

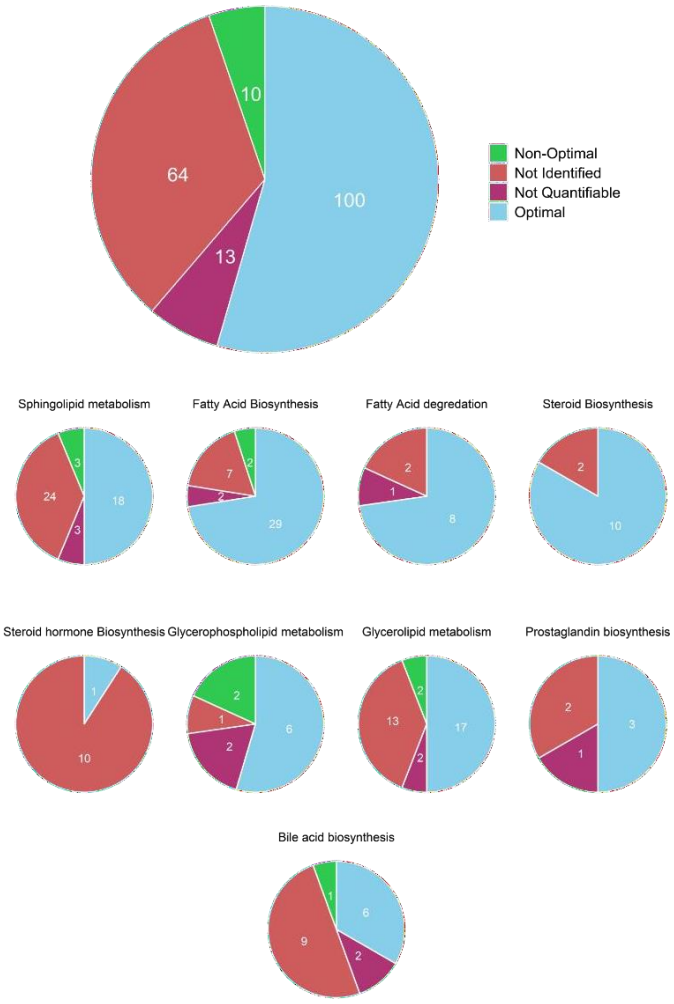

Fig8. Overall view of the progress in targeted method development. The pie chart shows the number of proteins that were not identified, proteins that were identified with no quantitypic peptides, proteins with only one quantitypic peptide, and proteins with more than one quantitypicpeptide.

Future Direction

The PRM assay is still missing many proteins from the initial list, and some of the identified proteins are still not measurable. The following step should be to identify the rest of the proteins and complete the method development step. We aim to identify theproteins using cDNA library-based in-vitro expression. Additionally, we should test the generated method and refine the targeting parameters to increase quantification accuracy. Finally, we will use heavy-labeled synthetic peptides to ensure the quality of our targeted

Reference List

1. H. Liu, R.G. Sadygov and J.R. YatesAnal. Chem., 76 (2004), pp. 4193-4201, 10.1021/ac0498563

2. S. Nahnsen, C. Bielow, K. Reinert and O. KohlbacherMol. Cell. Proteom., 12 (2013), pp. 549-556, 10.1074/mcp.R112.025163

3. S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb and R. AebersoldNat. Biotechnol., 17 (1999), pp. 994-999, 10.1038/13690

4. S.-E. OngMol. Cell. Proteom., 1 (2002), pp. 376-386, 10.1074/mcp.M200025-MCP200

5. P.L. Ross, Y.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan et al.Mol. Cell. Proteom., 3 (2004),pp. 1154-1169, 10.1074/mcp.M400129-MCP200

6. P. Picotti, B. Bodenmiller and R. AebersoldNat. Methods, 10 (2012), pp. 24-27, 10.1038/nmeth.2291

7. N. Selevsek, M. Matondo, M.S. Carbayo, R. Aebersold, B. Domon, Systematic quantification of peptides/proteins inurine using selected reaction monitoring, Proteomics 11 (2011) 1135e1147, http://dx.doi.org/10.1002/ pmic.201000599.

8. R. Huttenhain, M. Soste, N. Selevsek, H. Rost, A. Sethi, C. Carapito, et al., Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics, Sci. Transl. Med. 4 (2012) 142ra94, http://dx.doi.org/10.1126/scitranslmed.3003989

9. J.R. Whiteaker, C. Lin, J. Kennedy, L. Hou, M. Trute, I. Sokal, et al., A targeted proteomics based pipeline forverification of biomarkers in plasma, Nat.Biotechnol. 29 (2011) 625e634, http://dx.doi.org/10.1038/nbt.1900. V. Vidova, Z. Spacil / Analytica Chimica Acta 964 (2017) 7e2320

10. Marx, V., Targeted proteomics. Nat. Methods 2013, 10, 19-22.

11. Burgess, M. W.; Keshishian, H.; Mani, D. R.; Gillette, M. A.; Carr, S. A. Mol. Cell. Proteomics 2014, 13,1137−1149.

12. Kennedy, J. J.; Abbatiello, S. E.; Kim, K.; Yan, P.; Whiteaker, J. R.; Lin, C.; Kim, J. S.; Zhang, Y.; Wang, X.; Ivey, R. G.; Zhao, L.; Min, H.; Lee, Y.; Yu, M. H.; Yang, E. G.; Lee, C.; Wang, P.; Rodriguez, H.; Kim, Y.; Carr, S. A.;Paulovich, A. G. Nat. Methods 2014, 11, 149−155.

13. MacLean, B.; Bereman, M. S.; Ham, A.-J. L.; Schilling, B.; Thompson, J. W.; Dubois, L. G.; Held, J.

M..Cusack, M. P.; Bock, T.; Wollscheid, B.; Moseley, A.; Gibson, B. W.; Liebler, D. C.; MacCoss, M. J. 59th Annual ASMSConference on Mass Spectrometry, Denver, CO, June 5 - 9, 2011.

14. Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. Mol. Cell. Proteomics 2012, 11,1475−1488.

15. Sherrod, S. D.; Myers, M. V.; Li, M.; Myers, J. S.; Carpenter, K. L.; Maclean, B.; Maccoss, M. J.; Liebler, D. C.;Ham, A. J. J. Proteome Res. 2012, 11, 3467−3479.

16. Zhang C , Wang K , Yang L , Liu R , Chu Y , Qin X , Yang P , Yu H . Lipid metabolism in inflammation-relateddiseases. Analyst. 2018 Sep 24;143(19):4526-4536. doi: 10.1039/c8an01046c. PMID: 30128447.

17. Han X, Gross RW. The foundations and development of lipidomics. J Lipid Res. 2022 Feb;63(2):100164. doi:10.1016/j.jlr.2021.100164. Epub 2021 Dec 22. PMID: 34953866; PMCID: PMC8953652.

18. Kurz J, Parnham MJ, Geisslinger G, Schiffmann S. Ceramides as Novel Disease Biomarkers. Trends Mol Med.2019 Jan;25(1):20-32. doi: 10.1016/j.molmed.2018.10.009. Epub 2018 Nov 23. PMID: 30477968.

19. Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27-30(2000).

20. Kanehisa, M; Toward understanding the origin and evolution of cellular organisms. Protein Sci. 28, 1947-1951(2019)

21. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M.; KEGG: integrating viruses andcellular organisms. Nucleic Acids Res. 49, D545-D551 (2021).

22. Karlsson M et al., **A single-cell type transcriptomics map of human tissues.** *Sci Adv.* (2021)PubMed: 34321199 DOI: 10.1126/sciadv.abh2169

23. Cox, J., Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range massaccuracies and proteome-wide protein quantification. *Nat Biotechnol* **26,** 1367–1372 (2008). https://doi.org/10.1038/nbt.1511

24. Kong, A., Leprevost, F., Avtonomov, D. *et al.* MSFragger: ultrafast and comprehensive peptide identification inmass spectrometry-based proteomics. *Nat Methods* **14,** 513–520 (2017). https://doi.org/10.1038/nmeth.4256

25. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021 Jan8;49(D1):D480-D489. doi: 10.1093/nar/gkaa1100. PMID: 33237286; PMCID: PMC7778908.

26. Schilling B, Rardin MJ, MacLean BX, Zawadzka AM, Frewen BE, Cusack MP, Sorensen DJ, Bereman MS, Jing E, Wu CC, Verdin E, Kahn CR, Maccoss MJ, Gibson BW. Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. Mol Cell Proteomics. 2012 May;11(5):202-14. doi: 10.1074/mcp.M112.017707. Epub 2012 Mar 26.PMID: 22454539; PMCID: PMC3418851.