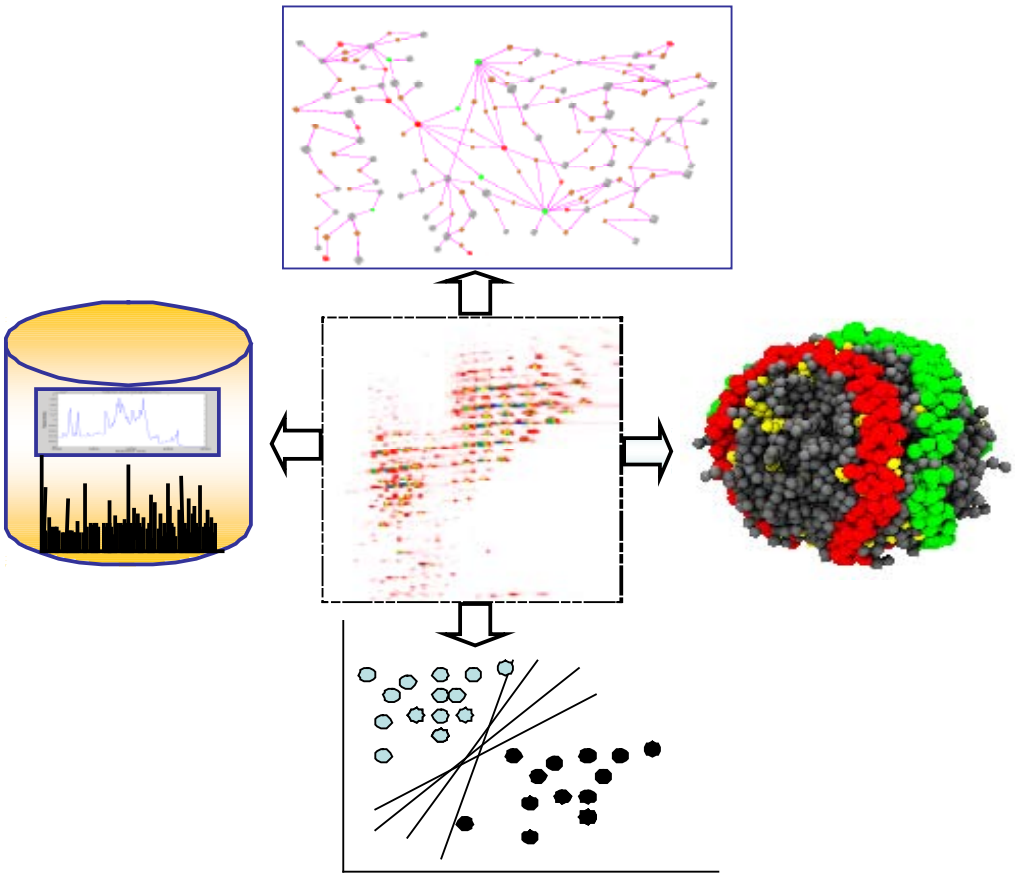
Dissertation VTT PUBLICATIONS 741



Laxmana Rao Yetukuri

Bioinformatics approaches for the analysis of lipidomics data

VTT PUBLICATIONS 741

Bioinformatics approaches for the analysis of lipidomics data

Laxmana Rao Yetukuri

Department of Biomedical Engineering and Computational Sciences (BECS)

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Faculty of Information and Natural Sciences, The Aalto University School of Science and Technology, for public examination and debate in Auditorium F239a at Aalto University (Otakaari 3, Espoo, Finland) on 11th of June, 2010 at 12 noon.

ISBN 978-951-38-7402-5 (soft back ed.)
ISSN 1235-0621 (soft back ed.)

ISBN 978-951-38-7403-2 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)

JULKAISIJA – UTGIVARE – PUBLISHER

VTT, Vuorimiehentie 5, PL 1000, 02044 VTT
puh. vaihde 020 722 111, faksi 020 722 4374

VTT, Bergsmansvägen 5, PB 1000, 02044 VTT
tel. växel 020 722 111, fax 020 722 4374

VTT Technical Research Centre of Finland, Vuorimiehentie 5, P.O. Box 1000, FI-02044 VTT, Finland
phone internat. +358 20 722 111, fax + 358 20 722 4374

Technical editing Maini Manninen

Edita Prima Oy, Helsinki 2010

Laxmana Rao Yetukuri. Bioinformatics approaches for the analysis of lipidomics data. Espoo 2010. VTT Publications 741. 75 p. + app. 106 p.

Keywords Lipids, Lipidomics, Bioinformatics, Lipid pathways, High density lipoproteins, k-nearest neighbours, Liquid chromatography/mass spectrometry, Principal component analysis, Partial least squares and discriminant analysis, Obesity, Support vector machines, LipidDB

Abstract

The potential impact of lipid research has been increasingly realised both in disease treatment and prevention. Recent advances in soft ionization mass spectrometry (MS) such as electrospray ionization (ESI) have permitted parallel monitoring of several hundreds of lipids in a single experiment and thus facilitated lipidomics level studies. These advances, however, pose a greater challenge for bioinformaticians to handle massive amounts of information-rich MS data from modern analytical instruments in order to understand complex functions of lipids**.** The main aims of this thesis were to 1) develop bioinformatics approaches for lipid identification based on ultra performance liquid chromatography coupled to mass spectrometry (UPLC/MS) data, 2) predict the functional annotations for unidentified lipids, 3) understand the *omics* data in the context of pathways and 4) apply existing chemometric methods for exploratory data analysis as well as biomarker discovery.

A bioinformatics strategy for the construction of lipid database for major classes of lipids is presented using simplified molecular input line entry system (SMILES) approach. The database was annotated with relevant information such as lipid names including short names, SMILES information, scores, molecular weight, monoisotopic mass, and isotope distribution. The database was tailored for UPLC/MS experiments by incorporating the information such as retention time range, adduct information and main fragments to screen for the potential lipids. This database information facilitated building experimental tandem mass spectrometry libraries for different biological tissues.

Non-targeted metabolomics screening is often get plagued by the presence of unknown peaks and thus present an additional challenge for data interpretation. Multiple supervised classification methods were employed and compared for the functional prediction of class labels for unidentified lipids to facilitate exploratory analysis further as well as ease the identification process. As lipidomics goes

beyond complete characterization of lipids, new strategies were developed to understand lipids in the context of pathways and thereby providing insights for the phenotype characterization. Chemometric methods such as principal component analysis (PCA) and partial least squares and discriminant analysis (PLS/DA) were utilised for exploratory analysis as well as biomarker discovery in the context of different disease phenotypes.

Preface

This thesis work has been carried out at Quantitative Biology and Bioinformatics (QBIX) group at VTT Technical Research Center of Finland (VTT). I derive great pleasure in thanking people who made this thesis a reality. Firstly, I would like to express my deep gratitude to my thesis advisor Research Professor Matej Orešič for providing this opportunity to pursue my PhD work in his fast growing QBIX group. The QBIX group gave me ample opportunity to work in the amazing field of lipidomics, which, in the beginning, was my unexplored territory of science. Prof. Orešič has been a constant source of encouragement, new ideas and inspiration throughout my PhD period. I would like to thank my co-advisor Dr. Jaakko Hollmén, Chief Research Scientist at Aalto University, for providing encouragement, scientific guidance and constructive criticism towards my thesis work. I wish to thank Prof. Kimmo Kaski who readily accepted me as his PhD student at Aalto University and followed the progress of my thesis. I greatly acknowledge my thesis pre-examiners Docent Pentti Somerharju and Prof. Seppo Auriola for their valuable comments which helped me present my thesis in a more meaningful way. I extend my sincere thanks to Technology Manager Dr. Richard Fagerström for his timely help in all walks of life at VTT. I am grateful to our former Vice President Prof. Juha Ahvenainen, Vice President Prof. Anu Kaukovirta-Norja and Research Professor Hans Söderlund for providing nice working facilities and environment for my research.

This thesis work would not have been possible without pleasant co-operation from my colleagues in our group. I thank the team leader of Biosystems Modeling, Dr. Marko Sysi-aho who encouraged and supported me to work on my thesis. I thank my close friend Mr. Gopalacharyulu Peddinti (Gopal) for his insightful scientific and non-scientific (heated) discussions. I like to thank Mr. Erno Lindfors and Mr. Mikko Katajamaa for developing software tools which I used for my research. I especially thank Erno, with whom I share my office place, for translating Finnish language documents for me.

I may be working in front of computer but data are coming from analytical lab. I greatly recognize the importance of all lab people who helped either directly or indirectly in generating the data. I thank the team leader of Metabolomics Dr. Tuulia Hyötyläinen for her continued support in this multi-disciplinary research. My special thanks go to Dr. Vidya Velagapudi and Dr. Tuulikki Seppänen-Laakso who generated most of the data on which this PhD is mainly based on. I thank Dr. Tapani Suortti and former team leader Dr. Kim Ekroos and Dr. Heli Nygren for broadening my analytical knowledge. I thank Dr. Mika Hilvo, Ms. Sandra Castillo, Mr. Pekka Savolahti, Mr. Matti Kankainen, Dr. Jing Tang, Dr. Catherine Bounsaythip, Dr. Perttu S. Niemelä and Mr. Han Zhao for their kind co-operation, friendly environment and scientific discussions. I wish to thank my previous supervisor at VTT, the team leader of Metabolic Engineering group, Dr. Laura Ruohonen for her guidance in yeast proteomics studies. I greatly acknowledge Dr. Marilyn Wiebe, Mrs. Eija Rintala and Dr. Laura Salusjärvi for their kind help during the proteomics related work. I also thank all my Indian friends who provided me social life outside of working place.

I express my gratitude to our collaborators who provided valuable samples for our research. In this regard, I thank Prof. Antonio Vidal-Puig, Prof. Marja-Riitta Taskinen, Prof. Matti Jauhiainen and Prof. Hannele Yki-Järvinen. I would not think of my PhD without their generous help of biological samples. I thank all my large list of coauthors for their kind co-operation in the publications.

Finally, I would like to take this opportunity to thank my family members. I like to thank my parents for their encouragement and blessings. I thank my wife, Krishnaveni Kannekanti, for her unconditional love and affection. I wish to express my warm greetings to my son, Bhuvan, for bringing new wave of happiness in my family.

List of publications

I L. Yetukuri, M. Katajamaa, G. Medina-Gomez, T. Seppänen-Laakso, A. Vidal-Puig, and M. Orešič (2007) Bioinformatics strategies for lipidomics analysis: characterization of obesity related hepatic steatosis, *BMC Syst. Biol.* 1: 12.

II L. Yetukuri, J. Tikka, J. Hollmén, and M. Orešič (2010) Functional prediction of unidentified lipids using supervised classifiers, *Metabolomics 6: 18–26.*

III L. Yetukuri, S. Söderlund, A. Koivuniemi, T. Seppänen-Laakso, P. S. Niemelä, M. Hyvönen, M.-R. Taskinen, I. Vattulainen, M. Jauhiainen, and M. Orešič (2010) Composition and lipid spatial distribution of High Density Lipoprotein particles in subjects with low and high HDL-cholesterol, *J. Lipid Res*. In press.

IV G. Medina-Gomez, L. Yetukuri, V. Velagapudi, M. Campbell, M. Blount, M. Jimenez-Linan, M. Ros, M. Orešič, and A. Vidal-Puig (2009) Adaptation and failure of pancreatic beta cells in murine models with different degrees of metabolic syndrome, *Dis. Model. Mech.* 2: 582–92.

V G. Medina-Gomez, S. L. Gray, L. Yetukuri, K. Shimomura, S. Virtue, M. Campbell, R. K. Curtis, M. Jimenez-Linan, M. Blount, G. S. H. Yeo, M. Lopez, T. Seppänen-Laakso, F. M. Ashcroft, M. Orešič, and A. Vidal-Puig (2007) PPARgamma 2 prevents lipotoxicity by controlling adipose tissue expandability and peripheral lipid metabolism, *PLoS Genet.* 3: e64.

VI A. Kotronen, V. R. Velagapudi, L. Yetukuri, J. Westerbacka, R. Bergholm, K. Ekroos, J. Makkonen, M.-R. Taskinen, M. Orešič, and H. Yki-Järvinen (2009) Serum saturated fatty acids containing triacylglycerols are better markers of insulin resistance than total serum triacylglycerol concentrations, *Diabetologia* 52: 684–690.

These articles are referred to as Publication I–VI in the text.

Author's contribution

I In Publication **I**, the author of this thesis involved in the research plan together with Matej Orešič to build the basic informatics work for lipidomics. The author constructed lipid database computationally, developed informatics methodology for lipid pathways, performed data analyses and drafted the manuscript. Mikko Katajamaa developed method for processing of UPLC/MS lipidomics data. Gema Medina-Gomez performed the experiments with ob/ob and WT animals. Tuulikki Seppänen-Laakso performed the lipidomics analysis. Antonio Vidal-Puig coordinated the *in vivo* studies and contributed to manuscript drafting. Matej Orešič initiated the study, performed data analyses and contributed to draft the manuscript.

II In Publication **II**, the author planned the study together with Matej Orešič and Jaakko Hollmén. The author performed programming tasks, carried out computational work and interpreted the results. The author drafted the manuscript. Jarkko Tikka participated in study design. Matej Orešič initiated study, helped in the study progress and provided valuable comments on the manuscript. Jaakko Hollmén supervised computational tasks and provided expert comments on the manuscript.

III In Publication **III**, the author carried out data preprocessing using MZmine software and identification of lipids. The author performed data analysis related to lipidomics and contributed to the major part of the manuscript drafting. Sanni Söderlund performed biochemical experiments and participated in data analysis. Artturi Koivuniemi performed molecular dynamics simulation studies and contributed to drafting of the manuscript, Tuulikki Seppänen-Laakso performed lipidomics experiments, Perttu S. Niemelä participated in discussions and co-ordinated modeling studies with Artturi Koivuniemi. Marja-Riitta Taskinen, Matti Jauhiainen, and Matej Orešič initiated and supervised the progress of the study. Perttu S. Niemelä, Marja

Hyvönen, Marja-Riitta Taskinen, Ilpo Vattulainen, Matti Jauhiainen, and Matej Orešič participated in the drafting and discussions of the manuscript.

IV In Publication **IV**, the author preprocessed the lipidomic data using MZmine software for all the tissues, and identified lipids. The author performed data analysis related to lipidomics and contributed to the manuscript drafting in statistical methods. Gema Medina-Gomez performed biochemical experiments in mouse, participated in discussion and the study design and drafted the manuscript. Vidya Velagapudi performed lipidomics experiments, Mark Campbell, Margaret Blount, Mercedes Jimenez-Linan and Manuel Ros participated in the biochemical experiments. Matej Orešič supervised the lipidomics studies and Antonio Vidal-Puig supervised the in vivo studies and drafted the manuscript.

V In Publication **V**, the author preprocessed the data with MZmine software and contributed to lipid identification and data analysis with Matej Orešič. Gema Medina-Gomez, Sarah L. Gray, Frances M. Ashcrof, and Antonio Vidal-Puig conceived and designed the experiments. Gema Medina-Gomez, Sarah L. Gray, Frances M. Ashcrof, Kenju Shimomura, Sam Virtue, Mark Campbell and Miguel Lopez performed the experiments. Gema Medina-Gomez, Sarah L. Gray, Kenju Shimomura, Sam Virtue, Mark Campbell, R. Keira Curtis, Mercedes Jimenez-Linan, Tuulikki Seppänen-Laakso, and Matej Orešič analysed the data. Gema Medina-Gomez, Kenju Shimomura, Sam Virtue, Mark Campbell, R. Keira Curtis, Margaret Blount, and Giles S. H. Yeo contributed reagents, materials and analysis tools. Gema Medina-Gomez and Antonio Vidal-Puig wrote the paper.

VI In the Publication **VI**, the author preprocessed the lipidomic data for all lipoprotein fractions. The author compiled tandem mass spectrometry (MS/MS)-based spectral library of all lipoproteins fractions combining both positive and negative ion mode. Anna Kotronen was involved in the data analysis and drafted the manuscript. Vidya R. Velagapudi performed lipidomics experiments and participated in the data analysis. Jukka Westerbacka, Robert Bergholm, Janne Makkonen were involved in the experimental design and biochemical experiments. Kim Ekroos was involved in the discussions on the manuscript. Marja-Riitta Taskinen, Matej Orešič and Hannele Yki-Järvinen were involved research design, discussions and drafted the manuscript. Matej Orešič participated in the data analysis.

Contents

Publications II and VI are not included in the PDF version. Please order the printed version to get the complete publication (http://www.vtt.fi/publications/index.jsp).

List of abbreviations

ANOVA	Analysis of variance
APCI	Atmospheric pressure chemical ionization
Cer	Ceramide
ChoE	Cholesteryl ester
ESI	Electrospray ionization
FDR	False discovery rate
HCA	Hierarchical clustering analysis
HDL	High density lipoproteins
HDL-C	High density lipoprotein-cholesterol
HPLC	High-performance liquid chromatography
KEGG	Kyoto encyclopedia of genes and genomes
k-NN	k-nearest neighbours
LC/MS	Liquid chromatography / mass spectrometry
lysoPC	Lysophosphatidylcholine
m/z	Mass-to-charge ratio

PC	Phosphatidylcholine
PC(e)	Ether linked phosphatidylcholine
PCA	Principal component analysis
POKO	Ablation of PPARg2 in Ob/Ob background (PPARg2(-/-) Lep(ob)/Lep(ob))
PE	Phosphatidylethanolamine
PE(p)	Ethanolamine plasmalogen
PLS/DA	Partial least squares and discriminant analysis
Q-TOF	Quadrapule time of flight
RT	Retention time
SM	Sphingomyelin
SMILES	Simplified molecular input line entry system
SVM	Support vector machines
TG	Triacylglycerol
TLC	Thin layer chromatography
UPLC/MS	Ultra performance liquid chromatography coupled to mass spectrometry

1. INTRODUCTION

Lipids are an important class of compounds that have a wide variety of key cellular functions including compartmentalisation, energy storage, cell-signalling, protein trafficking and membrane anchoring (Orešič et al. 2008, van Meer 2005, Vance and Vance 2008). Emerging evidence suggests that abnormalities in these functions are either directly or indirectly linked to the pathogenesis of various diseases (Wenk 2005) including obesity (Shi and Burn 2004), Alzheimer's diseas (Cutler et al. 2004), cancer (Menendez and Lupu 2007) and atherosclerosis (Lusis 2000). The lipids are generally hydrophobic in nature and are soluble in organic solvents. They are defined as hydrophobic or amphipathic small molecules that may originate entirely or in part by carbanion based condensation of thioesters, and/or by carbocation based condensation of isoprene units (Fahy et al. 2005). These lipids vary widely ranging from simple fatty acids to complex glycolipids. Systematic cataloguing of all these lipid classes and their nomenclature is required for databases and bioinformatics needs. According to new classification system, lipids are classified into eight main categories: 1) Fatty acids, 2) Glycerolipids, 3) Glycerophospholipids, 4) Sphingolipids, 5) Sterols, 6) Prenol lipids, 7) Saccharolipids, and 8) Polyketides (Fahy et al. 2009).

The structural diversity of lipids stems mainly from various combinations of fatty acid chain lengths and possible head groups (*e.g.*, for glycerophospholipids) that are linked to glycerol backbone (Figure 1.1). The presence of ethylene-interrupted or less common methylene-spaced double bonds in fatty acid moeities introduces an additional diversity in the lipidome. The presence of various types of glycerol-alkyl chain linkages such as ester, ether and vinyl ether bonds provide further diversity. Ether and vinyl ether bonds are more common in sn-1 position and are found mainly in phosphatidylcholine (PC) and phosphatidylethanolamine (PE) type of lipid classes (Snyder 1999). This diversity and abundance of these lipid species varies from tissue to tissue. For

instance, ether-linked (plasmanyl) and vinyl-ether linked (plasmenyl) phospholipids are more abundant in tissues such as heart, kidney, and central nervous system (Druilhet et al. 1975, Panganamala et al. 1971). Similar diversity can be seen in most other classes of lipids and the theoretical number of possible lipids, when conservatively estimated, exceeds ~180,000 lipids (Yetukuri et al. 2008).

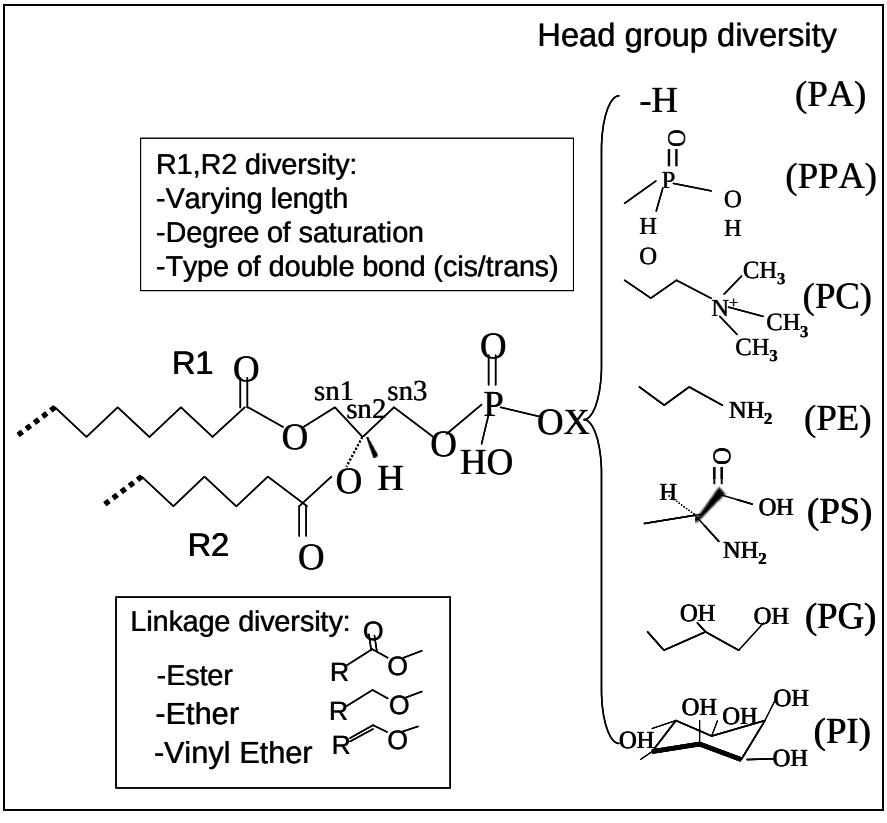


Figure 1.1. Glycerophospholipid structural diversity: R1 and R2 are fatty acids at position 1 and position 2, respectively. X represents head group moiety at position 3. Head groups legend: PA = Phosphate, PPA = Pyrophosphate, PE = Phosphoethanolamine, PC = Phosphocholine, PS = Phosphoserine, PG = Phosphoglycerol and PI = Phosphoinositol.

Recent burgeoning interest in lipid research illustrates the critical physiological importance of lipids. Moreover, recently emerged new lipidomics consortia such as the US-based LIPID MAPS (www.lipidmaps.org) and its affiliated sphinGOMAP (http://sphingolab.biology.gatech.edu/), and similar community-wide efforts in Japan (www.lipidbank.jp) and Europe (www.lipidomics.net) emphasis the growing need for indepth lipidomic research. Other related initiatives include The

Alliance for Cell Signaling (www.signaling-gateway.org/) and Lipid Profiles (www.lipidprofiles.com). As a result, lipid research is now beginning to appear as *omics* level science with the emerging precipitous developments. Lipidomics is defined as "the comprehensive understanding of the influence of all lipids on a biological system with respect to cell signaling, membrane architecture, transcriptional and translational modulation, cell-cell and cell-protein interactions, and response to environmental changes over time" (Watson 2006).

Any successful lipid analytical method should be a flawless combination of extraction, separation, detection and easy processing of the data. However, no single methodology or technique is yet in widespread use to screen all lipids. Recent advances in mass spectrometry-based studies have revolutionized lipidomics research at molecular level (Griffiths 2003, Navas-Iglesias et al. 2009, Ståhlman et al. 2009). Especially, the soft ionization MS-based analytical methods have gained popularity for their specificity, sensitivity and fast detection of different lipids from various biological matrices. The development of electrospray ionization techniques (Fernandis and Wenk 2009, Pulfer and Murphy 2003) have led to the study of lipids to a greater detail mainly in two ways: separation of lipid extracts using liquid chromatogram (LC) followed by on-line MS monitoring (Ogiso et al. 2008, Wang et al. 2005) and direct infusion of lipid extracts into a mass spectrometer where lipids are selectively detectable using techniques such as specific precursor ion scans (PIS) and neutral loss scans (NLS) (Ejsing et al. 2006, Ekroos et al. 2002, Han and Gross 2005a). As analytical technologies are becoming more mature and moving towards achieving the true quantitative or at least semi-quantitative characterization of molecular lipid species and lipid classes, it is becoming possible to study lipid pathways at the molecular level. It is evident that this knowledge will significantly advance our knowledge on thd roles of lipids in the context of cellular and organismal physiology.

Lipidomics, a branch of metabolomics, is the end point of *omics* cascade and bears the direct link to several disease phenotypes and therefore has recently become the target of post-genomics research. The advent of modern MS technologies has facilitated in the analysis of hundreds of lipid molecules from a given biological matrix. As a result, large scale data sets are being generated from the modern analytical methods, presenting new challenges for lipid informatics. The informatics approaches in co-ordination with improved analytical methods should resolve and identify individual lipids, unravel minute

systemic changes at molecular level and integrate the complex lipidome data with other enitities such as genes, proteins and other metabolites.

1.1 Aims of the thesis

The main aim of this thesis was to develop bioinformatics approaches for the non-targeted lipidomics data coming from UPLC/MS. The specific goals of my thesis (Figure 1.2) were the following:

- Development of computational spectral libraries and their customization for the UPLC/MS platform to facilitate lipid identification (Publication **I**).
- Building of tissue-specific experimental spectral libraries combining both positive and negative ion mode tandem mass spectrometry data (e.g., Publication **VI**).
- Functional class label annotation of unidentified lipids (Publication **II**).
- Mapping of lipidomics data in metabolic pathways (Publication **I**).
- Applications of developed informatics approaches followed by exploratory analysis (Publication **III, IV, V)**.

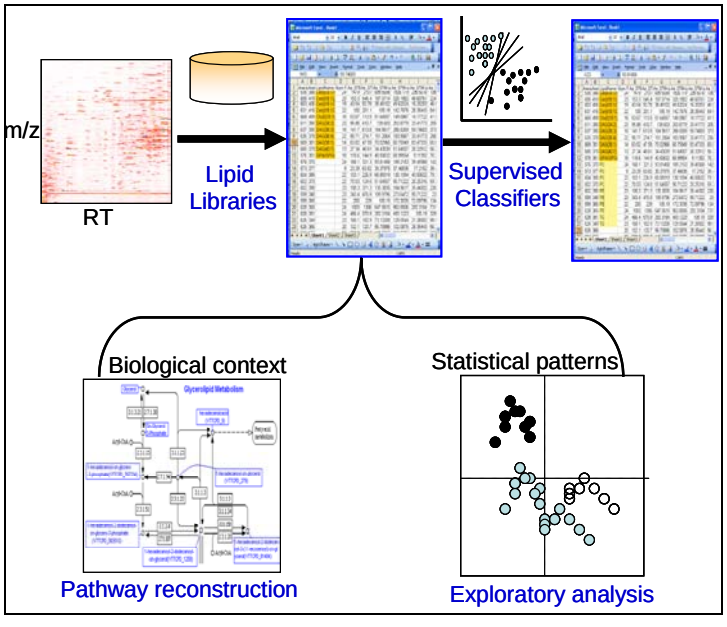


Figure 1.2. A schematic diagram showing the overview of this thesis in lipid bioinformatics approaches.

2. LITERATURE REVIEW

The lipid bioinformatics field is interdisciplinary in nature and topics related to this thesis are reviewed in this chapter. This chapter reviews the use of liquid chromatography and mass spectrometery in the lipid analysis, and advances in dataprocessing tools, databases and biochemical pathways.

2.1 Lipid analysis using LC/MS techniques

Biological matrix usually contains several lipids some of which have the same effective carbon number (ECN, a measure of non-polar characteristic of fatty acid chain or molecule) and therefore either co-elute partially or entirely from the liquid chromatography (LC) column. Thus simple LC alone is not enough to separate them. Favourably, mass spectrometry (MS) can act as second dimensional separation by discriminating compounds based on their mass fragments. Successful integration of LC with MS has played a significant role in the characterization of multiple compounds in a single sample.

2.1.1 Separation of lipids using liquid chromatography

Several studies have successfully demonstrated the usefulness of liquid chromatographic techniques in the analysis of complex lipid mixture (Bijlsma et al. 2005, Hermansson et al. 2005, Houjou et al. 2005). LC-based methods make use of properties such as differential solubility and partition between mobile and stationary phases. One of the earlier developments of LC is thin layer chromatography (TLC) which has been successfully used for the analysis of lipids (Bennet and Heftmann 1962, Michalec et al. 1962). TLC is developed from paper chromatography and comprises thin layer of stationary phase such as silica or cellulose on a flat support. Various combinations of aqueous stationary and organic mobile phases facilitate separation of several classes of lipids and

thus serve as a rapid screening tool prior to the development of more advanced and sensitive methods. Another notable methodology is solid phase extraction (SPE). This technology is useful in separating crude lipid mixtures into several classes including phospholipids, fatty acids, cholesterol esters, acyl glycerols and cholesterol (Kaluzny et al. 1985). High performance liquid chromatography (HPLC) has gained high popularity for its selectivity and efficiency. The separation in HPLC can be achieved using either normal-phase or reverse-phase columns. Normal-phase HPLC facilitates separation of lipids based on their polar head group (Lesnefsky et al. 2000) without major effect by the fatty acid substituents. In this case, each class co-elutes as single chromatographic peak with small retention time differences across the individual molecules in the class. In case of reverse phase column, the separation of lipids is based on their polarity, degree of alkyl chain saturation, and chain length. The reverse phase column thus enables separation of lipids with different fatty acid compositions (McHowat et al. 1997).

2.1.2 MS-based lipid analysis

Mass spectrometer, which has profound influence on modern analytical chemistry, can measure the mass of charge carrying molecules. The instrument measures both mass-to-charge ratio of molecule and its intensity and thus serves as an invaluable tool in structural elucidation (using tandem mass spectrometry) as well as quantification.

Electron ionization (EI) is useful in gas chromatography where the eluting gaseous molecules are bombarded with a beam of high-energy electrons and thus generating a specific fragmentation patterns. Other techniques such as chemical ionization (CI) use a reagent gas to ionize molecules which do not give molecular ion in EI and produce less fragmentation pattern than EI analysis. This ionization method was initially developed for gas chromatography and has been used as atmospheric pressure chemical ionization (APCI) for liquid samples. Laser-based soft ionisation technique called matrix-assisted laser desorption/ ionization (MALDI) is used for the analysis of large molecules and also can also be used for the analysis of lipids. The sample is mixed with a chemical matrix and then applied to sample holder as small spot. The matrix absorbs the energy from laser beam and thereby analyte receives the energy and results in ionization of molecules. ESI-based MS provides the most promising soft ionization technique and now has become the system of choice for both characterization

and quantification of lipids. This technology has been successfully applied in numerous lipidomic studies (Griffiths 2003, Han and Gross 2005a, Pulfer and Murphy 2003). This technique does not require any derivatisation and can directly be applied on solutions. The technique is moreover characterized by high-sensitivity with reasonable experimental complexity and high reproducibility. The ESI-based MS methodology has become the preferred method for the analysis of phospholipids (Brugger et al. 1997, Pulfer and Murphy 2003) and sphingolipids (Haynes et al. 2009). The methodology is also adopted for the analysis of non-polar lipids such as acylglyerols (Han and Gross 2001).

2.2 UPLC/MS-based non-targeted lipidomic analysis

Advances in MS-based analytical methods, in combination with ESI-based soft ionization technique, have spurred lipidomic research in recent years**.** The HPLC-based methods, however, are time-consuming and often present a bottleneck in the routine analysis. Shorter LC analysis times allow increasing sample throughput as well as lab productivity.

Recent advances in novel mass analysers such as orthogonal-accelerated time of flight (oa-TOF) have led to the development of mass spectrometer. A commonly used mass spectrometer with oa-TOF analyser contains a quadruple-time of flight (Q-TOF) configuration (Figure 2.1) and resembles as triple quadrupole MS in which third quadrupole has been replaced with oa-TOF. The TOF analyser permits full scan acquisitions with high resolution and mass accuracy. This powerful combination of single quadrupole and high performance of TOF enables both MS and tandem mass spectrometry (MS/MS) experiments. In the MS mode, the first quadrupole and collision cell merely guide the ions, while TOF seperates all the ions that are orthogonally accelerated. In MS/MS mode, the filter capabilities of quadrupole are exploited to transmit and to select only precursor ions. These precursor ions are accelerated due to potential difference before they get fragmented in the collision cell induced by collision with neutral gases such as argon or nitrogen. The resulting ions are analysed in TOF tube with high mass accuracy.

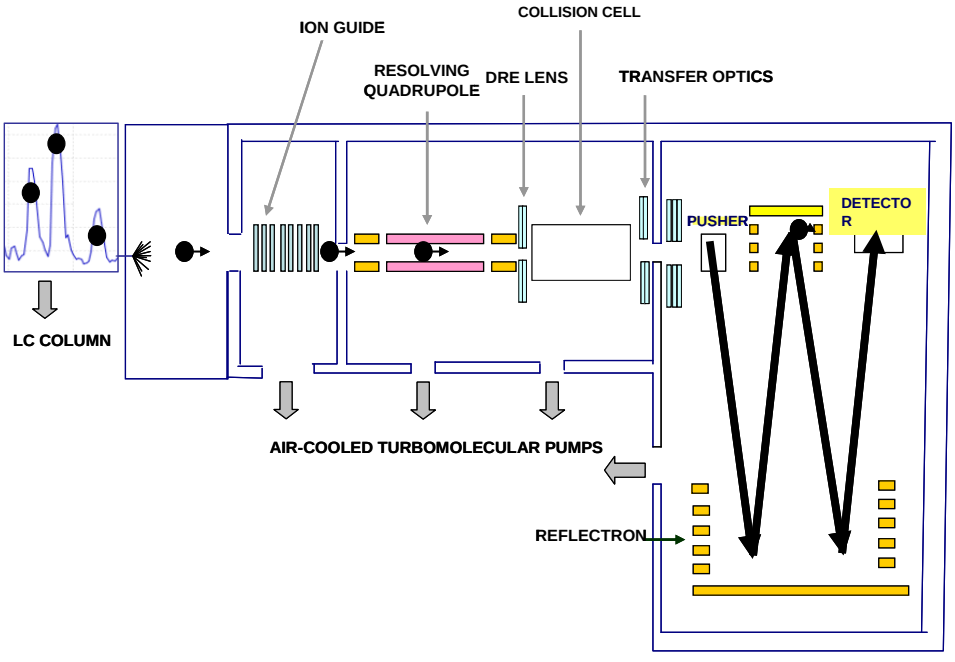


Figure 2.1. Schematic view showing the configuration of quadrupole time of flight (Q-TOF) mass analyzer.

Recent advances in reverse phase chromatographic columns (e.g., Bridged Ethane Hybrid (BEH) C18 columns packed with 1.7 µm particles) and mobile phase systems allowed HPLC system to operate at much higher back-pressures. The new column, called ultra performance liquid chromatography (UPLC), offers significant advantages in resolution, speed, and sensitivity as compared to conventional HPLC analysis (Apollonio et al. 2006, Churchwell et al. 2005, Leandro et al. 2006, Wilson et al. 2005). With improved speed and sensitivity, UPLC/MS platform thus provide a greater advantage in high-throughput sample analysis. Here, UPLC is coupled to high resolution quadrapole time of flight (Q-TOF) mass analyzer which enables accurate mass measurements of precursor and fragment ions. Non-targeted lipidomics approaches have the advantage of detecting greater number of lipid components and possibility for detecting novel compounds. It is also advantageous in picking up global changes thereby serving as a guide for designing targeted approaches.

2.3 Data processing tools for lipid analysis

The modern analytical instruments allowed lipidomic studies, generating huge amounts of data to process. The data processing step is critical, labor-intensive and becomes the rate-limiting step in metabolomics studies. This impetus has led to the development of many data processing tools both in open source and commercial environment (Katajamaa and Orešič 2007). Common data processing steps in lipidomics include peak detection, lipid identification, isotope correction, response correction and quantification. The typical data processing starts by inputting data in some vendor-specific format, which often becomes practical difficulty in applying same software for different data formats coming from different vendors. Recently introduced tools such as Lipid Qualitative/Quantitative Analysis (LipidQA) software platform (Song et al. 2007) can identify and quantitate the complex lipids in biological mixtures. The software can process the data coming from TSQ-7000 triple stage quadrupole and LTQ linear ion trap mass spectrometers from Thermo-Finnigan and Q-TOF hybrid quadrupole/time-of-flight instrument from Waters-Micromass. The algorithm can also handle data-dependent manner lipid identification based on MS/MS spectra of glycerophospholipid species. Fatty Acid Analysis Tool (FAAT) tool (Leavell and Leary 2006) is developed for the analysis of data coming from Fourier transform mass spectrometry and the tool is demonstrated using mycobacteria species data. The main functionalities of the software include identification of overlapping saturated and unsaturated lipids, assignment of known ions from a user-defined library and handling of isotopic shifts from stable isotope labeling experiments. The software tools such as SECD and LIMSA (Hermansson et al. 2005) are useful for the display of chromatograms and performing several data processing steps including peak picking, integration, isotope correction and internal standards-based quantification. Other tools such as Lipid Profiler (Ejsing et al. 2006) and LipidInspector (Schwudke et al. 2005) are compatible for the data acquisition with Applied Biosystems hybrid quadrupole/time-of-flight instruments that can perform multiple precursor ion scans in a single experiment. The novel lipid-mediator informatics developments such as cognoscitive-contrast-angle algorithm and database (COCAD) (Lu et al. 2006) enhance correct identifications of lipid-mediators by matching either known standard MS/MS spectra with chromatograms and UV spectra or virtual liquid chromatography-ultraviolet-tandem mass spectra.

MZmine software (Katajamaa et al. 2006, Katajamaa and Orešič 2005) is an open source Java-based data processing tool for LC/MS-based metabolomics experiments, with a particular focus on differential analysis of lipidomics data from UPLC/MS experimental setup. The software implements several key methods for data processing stage including spectral filtering, peak picking, deisotoping, alignment of samples and quantification. Moreover, recursive peak search algorithm and peak picking methods facilitate the improvement of already aligned data. Several data visualization options are available for the display of spectral data across multiple samples. MZmine2 (http://mzmine.sourceforge.net/) presents new features and improved modularity for better expandability. Some of new features include processing of high-resolution instrumental data, better visualisation (3D visualisation) and a new implementation of 2D visualiser. It allows storage of parameters for defining sample properties. The software supports importing of several data formats such as netCDF, Thermo RAW, mzML and mzXML and also the stored project-specific parameters defining the samples.

2.4 Lipid databases

Development of databases (Table 2.1) and related bioinformatics tools has become an essential part of functional genomics studies. Over the recent years, empowered by high-throughput technologies for *omics* fields, creation of databases devoted to certain entities such as lipids was undertaken. Consequently, lipid-centric databases were developed that enabled researchers to comfortably analyse expression patterns of lipid related genes and gene products. For example, a database of genomics of lipid-associated disorders, called GOLD, offers annotated pathways, curated data sets and possibility to study experimental data in the context of biological pathways (Hackl et al. 2004). The LIPID MAPS Proteome Database (LMPD) (Cotter et al. 2006) is a database of lipid-associated protein sequences and annotations. Presently, the database mainly comprises human and mouse related proteins of lipid metabolism. The protein database is enhanced with annotations from external databases.

Several lipid databases such as LIPID BANK (http://www.lipidbank.jp/), LIPIDAT (Caffrey and Hogan 1992) and LMSD (Sud et al. 2007) are publicly available offering wide-range of information including lipid structures. Notably, LMSD offers systematic structures of lipids as well as other related information according to the classification scheme recommended by LIPID MAPS consortium. Users can retrieve the data from LMSD using text- or structure-based queries.

Web tools such as LIPID MAPS online tools (Fahy et al. 2007) permit user friendly-queries from underlying databases. Lipid library (http://www.lipidlibrary.co.uk/) and CyberLipid (http://www.cyberlipid.org/) offer rich source of lipid information.

The proposed classification, nomenclature, and chemical representation system introduced by the the LIPID MAPS consortium has become standard reference for the construction of lipid databases and management of lipidomics data. One of the main goals of LIPID MAPS initiative includes building of lipid databases and related bioinformatics approaches. These databases need to be complemented with annotation and curation of lipid structures. The LIPID BANK aims at curation of lipid structures as well as annotation of the related literature. Other databases like PubChem (http://pubchem.ncbi.nlm.nih.gov/) offer huge repository of chemical compounds including lipids. The database also provides various physical/chemical properties and cross links to other databases. All these databases are handy in the analysis of lipids and may serve as tools for validation of results. Notably, LMSD also provide various tools for mass spectrometry data. However, given the diversity of lipids across different organisms, tissues, and cell types, it is unlikely any one database can become a reference for mass spectrometry data. Mainly for LC/MS-based analysis, the development of in-house databases that are customized for the instrumental settings are needed

2.5 Lipid pathway resources

The existing databases offer rich source of information on lipid pathways. Databases such as Kyoto encyclopedia of genes and genomes (KEGG) database (Kanehisa and Goto 2000, Kanehisa et al. 2004) serve as a valuable resource for analyzing cells, not only at genomic level but also for metabolic networks in different organisms. The database offers information on most metabolic pathways including lipid pathways. Additionally, KEGG provides generic pathways (i.e., species-independent pathways) as reference pathways for the reconstruction of context- or organism-specific pathways. Moreover, the KEGG Brite (http://www.genome.jp/kegg/brite.html) maintains a collection of hierarchical classifications of lipid species whose reactions and pathways can be viewed. Other more annotated databases such as MetaCyc (Krieger et al. 2004) and EcoCyc (Keseler et al. 2005) serve as a good starting point for the study of lipids. SphinGOMAP (http://sphingolab.biology.gatech.edu/) offers comprehensive pathway mapping of about 450 distinct sphingolipids and glycosphingolipids species. LIPID

MAPS biopathways workbench (http://www.biopathwaysworkbench.org/) provides a graphic tool that facilitates to display, edit and analyse biochemical pathways of lipids**.** In a recent study (Gupta et al. 2009), kinetic model was built from the lipidomics flux analysis using integrated network of eicosanoids metabolism and signaling pathways. The developed integrated model is based on the KEGG pathways and literature knowledge. Rate constants in the kinetic model are estimated and tuned using generalised constrained non-linear optimization. These quantitative models are quite useful for perturbation studies to gain mechanistic understanding about the underlying phenotype. These tools and databases allow reconstruction of integrated pathway models and thus open new avenues for building system level quantitative models.

Table 2.1. List of publicly available lipid resources and their characteristics.

Lipid database/source	**Description**
LIPID MAPS (www.lipidmaps.org)	Provides guidelines for classification system for lipids, hosts databases of lipids and lipid-associated protein data and develop tools for identifying lipids.
Lipidomics Expertise Platform (http://www.lipidomics.net/)	European level Initiative for lipid research. Provides databases for the registered users and serves as a source for establishing European level networks, industrial relations and exchange of standard materials.
CyberLipids (www.cyberlipid.org)	Offers huge collection of updated scientific knowledge on all aspects of lipids. Also facilitates establishing relationships among students, teachers, scientists and technicians and present and provides updated bibliography devoted to lipid biology.
LIPIDAT (ww.lipidat.chemystry.ohio-state.edu/home.stm)	Presents thermodynamic information on lipids including lipid phase transition temperatures and enthalpy changes for synthetic and biologically relevant complex polar lipids.
LIPID BANK (www.lipidbank.jp)	Provides chemical structures of lipids with names, chemical and physical properties, biological activities and metabolism. In addition, spectral information from various instruments such as ultraviolet, infrared spectrometry, nuclear magnetic resonance, mass spectrometry, liquid chromatography, and thin-layer chromatography can also be obtained.
KEGG lipids (http://www.genome.jp/kegg-in/get_htext?br08002.kegg)	Provides lipid pathway maps as well as associated information such as the name, formula, mass, structure, biochemical reactions and external links to other public databases
THE LIPID LIBRARY (http://www.lipidlibrary.co.uk/)	Portal for the study of many classes of lipids and their analysis both in mass spectrometry and chromatography.
sphinGOMAP (http://www.sphingomap.org/)	Offers a database for biochemical mapping of sphingo- and glycosphingo-lipids.

3. METHODS

3.1 Lipid analysis using UPLC/MS platform

3.1.1 Lipid standards and chemicals

Sample was extracted with chloroform / methanol (2:1, 100 µl) after addition of an aliquot (20 µl) containing internal standard mixture. After homogenization and vortexing, the sample was centrifuged (10000 rpm, 3 min) and the lower phase was collected. The lower lipid phase was mixed with another aliquot of labeled standard mixture. The labeled mixture containing 3 standards was added before analysis in order to control the extraction process (Pietiläinen et al. 2007). The labeled standards were PC(16:0/0:0-d3), PC(16:0/16:0-d6) and TG(16:0/16:0/16:0-^{13}C3) and were obtained from Larodan Fine Chemicals (Malmo, Sweden). The internal standards mixture comprised MG(17:0/0:0/0:0)[rac], DG(17:0/17:0/0:0)[rac] and TG(17:0/17:0/17:0) from Larodan Fine Chemicals (Malmo, Sweden) and PC(17:0/0:0), PC(17:0/17:0), PE(17:0/17:0), PG(17:0/17:0)[rac], Cer(d18:1/17:0), PS(17:0/17:0), PA(17:0/17:0) and D-erythro-Sphingosine-1-Phosphate (C17 Base) from Avanti Polar Lipids (Alabaster, AL).

3.1.2 Mass spectrometry conditions

The extracted lipid samples were analysed on quadrupole time-of-flight (Q-Tof Premier) mass spectrometer combined with an Acquity ultra performance liquid chormatogram (UPLC) (Waters Inc., Milford, MA). The column was an Acquity UPLC™ BEH C18 10×50 mm with particle size of 1.7 µm and was maintained at 50°C. The composition of the binary solvent system at the flow rate of 0.200 ml/min was A: water (1% 1 M NH_4Ac, 0.1% HCOOH) and B: LC/MS grade (Rathburn) acetonitrile/isopropanol (5 2, 1% 1 M NH_4Ac, 0.1% HCOOH). The

initial gradient of the solvent composition was 65% A/35% B. The gradient reached 100% B in 6 min and maintained there for the next 7 min. The total run time was 18 min which included a 5 min re-equilibration step. The sample organizer was set at 10°C.

The lipid profiling was carried out on Waters Q-Tof Premier mass spectrometer using electron spray ionization in either positive or negative ion mode. The data were collected usually in the mass range of m/z 300–1200 with scan duration of 0.2 sec. The temperature of source was maintained at 120°C and nitrogen was used as desolvation gas (800 L/h) at 250°C. The voltages of the sampling cone and the capillary were 39 V and 3.2 kV, respectively. Reserpine (50 µg/L) was used as the lock spray reference compound (5 µl/min; 10 sec scan frequency). The samples were analysed in a randomized order.

3.1.3 Data processing with MZMine

Lipid data from UPLC/MS experiments were first converted from raw data to netCDF file format using DataBridge utility of MassLynx 4.1 software (Waters, Inc.). The netCDF files were preprocessed using an in-house developed MZmine software version 0.60 (Katajamaa et al. 2006). Main functionalities of the software include peak picking, chromatographic alignment, spectral filtering, peak area calculations, visualisation (i.e. peak maps, curvilinear distance analysis and Sammon's mapping), gapfilling, normalisation, and data export. De-isotoping step was performed using in-house developed MATLAB scripts. Lipids were identified using an internal spectral library or alternatively with tandem mass spectrometry. Calibration (normalisation) was done based on multiple internal standards and was performed as follows: All monoacyl lipids (monoacylglycerols and lysophospholipids) were normalized with LysoPC(17:0/0:0), all diacyl lipids except phosphatidylethanolamines and ethanolamine plasmalogens were normalized with PC(17:0/17:0), the phosphatidylethanolamines and ethanolamine plasmalogens were normalized with PE(17:0/17:0), and the triacylglycerols and cholesterol esters with TG(17:0/17:0/17:0). Calibration of unidentified lipids, similar to method described earlier (Bijlsma et al. 2005), was done using three internal standards as follows: lysoPC(17:0/0:0) was used to normalize the peaks eluting with retention time (RT) < 300s, PC(17:0/17:0) for 300s < RT< 410s, and TG(17:0/17:0/17:0) for RT > 410s.

3.2 Lipid database construction

3.2.1 Lipid scaffold generation

Lipid database was constructed computationally using Simplified Molecular Input Line Entry System (SMILES) approach (Publication **I**) which is a widely used chemical notation to represent a chemical structure in terms of atoms and bonds governed by set of syntax rules (Weininger 1988). The database accommodated main classes of lipids including fatty acids, phospholipids, glycerolipids, cholesterol esters, and sphingolipids. The scaffolds of theoretically possible lipids were computed based on known lipid building blocks such as polar head groups and fatty acids in order to facilitate identification of lipids. In order to construct a particular lipid class (e.g., glycerophospholipid), SMILES template was constructed to represent the structure of the class. Possible lipids in the class along with their names were generated compuationally by varying alkyl moiety, nature of linkage and head group. Each lipid in the database was annotated with systematic name, SMILES, molecular formula and exact average mass and monoisotopic mass. A score value was assigned to each compound based on natural abundance of fatty acid(s). Common factors considered while assigning the score were natural abundance of the fatty acid, and odd or even number of carbon atoms present in the fatty acid chain. This kind of scoring scheme facilitated quick search of possibly abundant compounds.

3.2.2 Lipid nomenclature

Lipids were named according to recent nomenclature system for lipids introduced by LIPID MAPS consortium (Fahy et al. 2009). For example, lysophosphatidylcholine with 17:0 fatty acid chain at sn-1 position was named as 1-heptadecanoyl-sn-glycero-3-phosphocholine (short name: PC(17:0/0:0)). If the exact fatty acid composition was not determined, total number of carbons and double bonds was indicated. For example, a phosphatidylcholine species PC(18:0/20:4) is represented as PC(38:4). However, PC(38:4) may correspond to isobaric (e.g., PC(22:4/16:0)) or isomeric species (e.g., PC (20:4/18:0)).

3.2.3 Database design

Information on each lipid entry was converted into XML document and the resulting documents were loaded to the database using mass-loading tool of Tamino server. The database is a native XML database implemented in Tamino XML Server (Software AG). Each entry in the database was annotated with an internal identifier, scoring information, class, canonical SMILES, molecular formula, molecular weight and isotopic distribution. All the relevant scripts were implemented in the Perl language. When implementing the database, we used XMLSPY software (Altova, Inc.) and Tamino Schema Editor Software (Software AG) for the construction and validation of logical and physical schemas, respectively.

3.3 Lipid pathways

3.3.1 Lipid pathways and extensions

System level characterization by integrating genes, proteins, lipids and other molecules provide better insights in an organism (Joyce and Palsson 2006). The existing databases provide different levels of information. The databases such as KEGG database (Kanehisa and Goto 2000, Kanehisa et al. 2004) serve as a valuable resource for analyzing cells not only at genomic level but also for metabolic networks in different organisms. The KEGG is a database of biological systems that integrates genomic, chemical and network information (http://www.genome.jp/kegg/). The KEGG hosts a collection of manually drawn pathway maps based on the current knowledge on the molecular interaction and reaction networks. The KEGG PATHWAY database offers information on most metabolic pathways including lipid pathways include fatty acid biosynthesis, fatty acid elongation in mitochondria, fatty acid metabolism, synthesis and degradation of ketone bodies, steroid biosynthesis, primary and secondary bile acid biosynthesis, C21-Steroid hormone metabolism, androgen and estrogen metabolism, glycerolipid metabolism, glycerophospholipid metabolism, ether lipid metabolism, sphingolipid metabolism, arachidonic acid metabolism, linoleic acid metabolism, alpha-linolenic acid metabolism and biosynthesis of unsaturated fatty acids. Additionally, KEGG also provides generic pathways (i.e., species-independent pathways) to serve as reference pathways for the reconstruction of context- or organism-specific pathways.

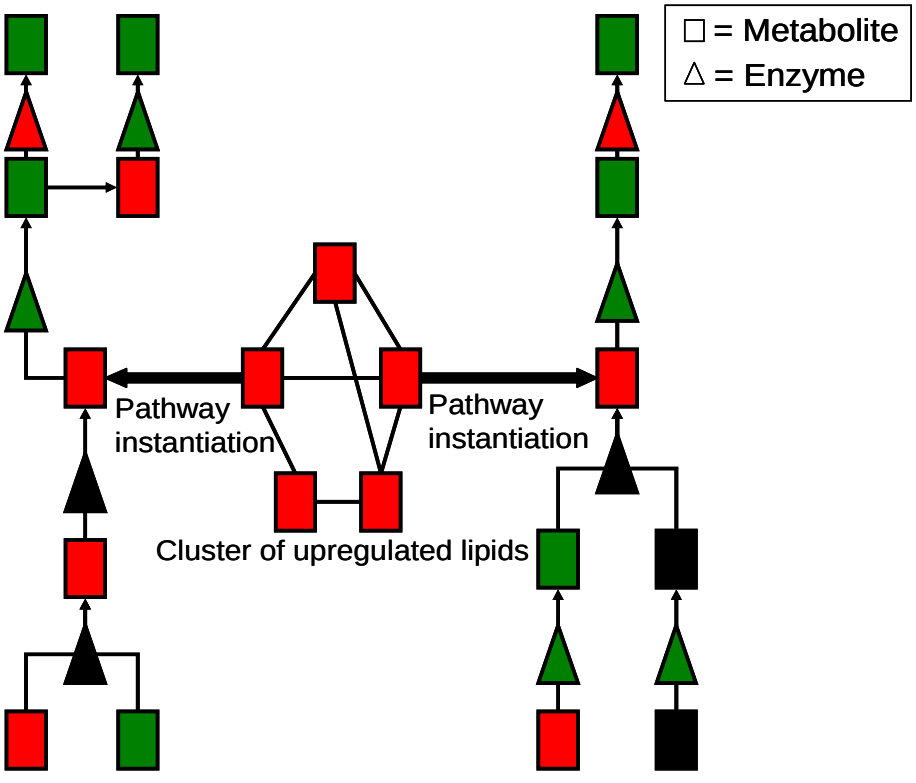


Figure 3.1. An illustrative example of lipid pathway reconstruction methodology via pathway instantiation. Interesting lipids from co-regulation network are linked to molecular instance pathways. Red/green colour coding refers to up/down regulation of enzymes and lipids.

Modern MS-based analytical techniques provide lipid species level information, whereas present lipid pathway information is mostly limited to the generic lipid class level. At the generic level, most lipid entries on pathways may contain one or more fatty acids and/or head groups. Due to enormous diversity in fatty acids and head groups, large number of specific lipids that are measurable now can be substitutable for a particular entry on pathway. As a result, lipid pathway reconstruction may easily end up in combinatorial explosion with varying complexity from pathway to pathway. To avoid this combinatorial problem, generic pathway templates are utilized to create molecular instance pathways for molecular species selected based upon multivariate and co-regulation analyses (Figure 3.1). Pathway instantiation is a method of converting generic names of lipids on biochemical lipid pathways to corresponding specific names of interest. Instance pathways allows mapping of lipids observed in mass spectrometric experiments.

3.3.2 Visualization tools for lipid pathways

Managing and organising lipid-related pathways into useful, interactive pathways and networks present a greater challenge for lipid bioinformatics. In-house developed megNET software tool enables retrieval and visualisation of biological relationships across heterogeneous data sources from an integrated database (Gopalacharyulu et al. 2005). Other open source visualization tools such as VANTED (Junker et al. 2006) enable importing and customisation of KEGG lipid-specific pathways.

3.4 Modeling with supervised and unsupervised methods

There are two main categories of mathematical modeling approaches: supervised and unsupervised. Supervised modeling (Mitchell 1997) is an approach that uses pairs of input objects (usually in the form of matrix, **X**) and desired outputs (usually in the form of matrix, **Y**). The output of the function can either be continuous, as in the case of regression, or categorical, as in the case of classification. Information in matrix **Y** is used to guide the construction of the model for **X** and hence the name supervised. On the other hand, unsupervised modeling only utilizes the observed data in X and the model tries to learn the statistical patterns or trends available in X (Duda et al. 2001). Conventional methods for unsupervised learning such as principal component analysis (PCA) and hierarchical cluster analysis (HCA) are generally employed in exploratory analysis.

3.4.1 Preprocessing of multivariate data

Preprocessing of multivariate data is advocated to extract relevant information from a given data matrix. One of the most commonly employed procedures is mean-centering (also called column centering) where the goal is to model the actual variation in the data. In mean-centering, the mean of each measured variable (column mean) is substracted from each value of the respective variables in the data set so that resulting data matrix contains columns with zero mean. Other routinely used preprocessing step, especially in chemometrics, is unit variance (UV) scaling where each variable (column vector in the data matrix) is divided by the respective standard deviation. This scaling alleviates the effect of differences in magnitude of variables i.e., the higher magnitude

variables have the greater influence on the results. Combination of mean-centering and UV-scaling can also be employed (i.e., auto scaling).

3.4.2 PCA

Principal Component Analysis (PCA) is a latent variables-based unsupervised method for exploratory analysis (Hotelling 1933). PCA uncovers simpler patterns from the complex inter-correlated variables. The PCA can also be seen as a dimensionality reduction strategy while retaining as much information as possible. This is achieved by creating new set of variables which are linear combinations of original variables to produce principal components. These principal components are orthogonal to each other and are uncorrelated. The first principal component is in the direction of the greatest variance in the data and subsequent components are constructed orthogonal (independent) to the previous ones in the direction of largest remaining variance. Usually, the first few latent components account for the most of the variation in the data matrix (X [nxm]). The data matrix, X, can be decomposed into two matrices: scores matrix (T) and loadings matrix (P)

$$X = TP^t = \sum_{i=1}^{m} t_i p_i^t \tag{4.1}$$

where t denotes a transpose operation on matrix. The loading matrix contains information on the variables while the scoring matrix contains information about the objects. When the data are projected into a lower dimensional space spanned by few principal components corresponding to maximum variation, the data matrix X can be written as in equation (4.2).

$$X = \sum_{i=1}^{k} t_i p_i^t + E \tag{4.2}$$

where E is unexplained variation in data matrix X and k (k << m) is the number of first principal components.

3.4.3 HCA

General clustering methods are based on the distance between the samples whose observed parameters are co-ordinates in the multi-dimensional space.

Similarity or dissimilarity of the samples is based on whether they are close or not in the high dimensional space. Hierarchical cluster analysis (HCA) (Jolliffe 1986) can broadly be divided into two methods: agglomerative methods and divisive methods. The divisive methods start with all of the observations in one cluster and then proceed to split (partition) them into smaller clusters. On the other hand, the agglomerative methods initially treat each observation as a separate cluster and then proceed to fuse pair of clusters with smallest distance. The fusion will continue until all observations belong to one cluster. Popular agglomerative methods are single linkage (nearest neighbor approach), average linkage, complete linkage (furthest neighbor) and Ward's method.

Average linkage clustering uses the average similarity of observations between two groups as the measure between the two groups. Complete linkage clustering uses the furthest pair of observations between two groups to determine the similarity of the two groups. Single linkage clustering, on the other hand, computes the similarity between two groups as the similarity of the closest pair of observations between the two groups. Ward's linkage is distinct from all the other methods in that it uses an analysis of variance approach to evaluate the distances between clusters.

The outcome of HCA is a hierarchy or tree-like structure (dendogram) showing the relations among the entities. Dendogram can be interpreted based on the length of branches which are proportional to distance between various clusters.

3.4.4 PLS-DA

PLS-DA is a widely used supervised classification algorithm when dimensionality reduction is needed and discrimination is sought in multivariate analysis (Matthew Barker 2003). In the mass spectrometry data, it is typical to observe that the number variables are more than the number of samples. Moreover, many variables are correlated. The partial least squares (PLS) method permits investigation of complex problem of collinearity (i.e., X-variables). The PLS-DA model establishes the relation between predictor variables (i.e., X matrix) and response variables (i.e., Y matrix) by finding latent variables in such a way that the covariance between the two variables is maximum. The obtained latent variables are linear combinations of old X-variables. Often optimum number of latent variables is needed and can reliably be computed from cross validation procedures.

3.4.5 k-NN

k-NN is a supervised learning algorithm that works without any prior assumptions about the distribution from which training samples are drawn. The training data are vectors in the high-dimensional feature space which is partitioned into regions by class label. The algorithm involves storing of feature vectors and class labels in the learning phase. In order to estimate the class label for the test sample, k-NN computes distances from test sample to all other samples from training set and k nearest neighbours are selected. In order to compute the distances, distance metrics such as Euclidean distance can be employed. Test sample is assigned to most frequent class label among k nearest neighbours in high dimensional feature space. If the tie scenario arises, the ties are broken at random or closer neighbours are given priority. It is common to select larger k value to help reduce noisy effects in data and odd k value to break ties. The optimal choice of k is important and can be selected by cross-validation approach (Duda et al. 2001).

3.4.6 SVM

Support vector machines are a group of supervised methods introduced earlier by Vapnik (Vapnik 1995). The SVMs have gained popularity and have been successfully applied to number of applications including protein structural classification, image recognition, text classification, microarray gene expression data analysis and protein fold recognition (Brown et al. 2000, Cai et al. 2001, Joachims 1998). SVMs are primarily designed for binary classification problems where the training data with two classes are transformed into a high dimensional space by kernel functions. These classifiers rely on hyperplanes corresponding to decision functions. SVM model achieves its objective of classification by constructing optimal hyperplane, i.e., the hyperplane that maximises separation between the two classes. The solution for the classification lies in the support vectors that determine the maximum margin hyperplane. The margin of a linear classifier is the minimal distance of any training point to the hyperplane. Multi-class problem of SVMs can be regarded as multiple binary class problems. One way to solve multi-class classification is using ‘one-versus-one’ approach where the model constructs a binary classifier for every pair of classes, resulting in k (k-1)/2 SVM models for k-class classification problem.

3.4.7 Naive Bayes

Naive Bayes is a probability based classifier and is obtained by assuming conditional independence of the predictor variables given the class label in Bayes theorem (Hand et al. 2001). As a result, likelihood term of Bayes rules can be decomposed into product terms. The classifier basically ignores the potential inter-dependencies such as correlations among the inputs and reduces a complex multivariate problem to a group of simple univariate problems. For a given set of predictor variables, X, the model constructs posterior probability for any event C_j among the set of categorical levels of C. Labeling of new predictor variable to a particular class is based on the highest posterior probability. Naive Bayes methodology simplifies a classification task by allowing the computation of class conditional densities for each variable separately.

3.5 Statistical hypothesis testing

3.5.1 Student's t-test

Two sample t-test (Fisher Box 1987, Snedecor and Cochran 1989) is commonly employed to investigate whether the means of two groups of samples are significantly different from each other. The t-test compares difference in the two means in relation to existing variation in the data. The t-test, as shown in equation (4.3), is a ratio with numerator representing the differences between the means and denominator denoting the measure of variability (standard error of differences) in the data.

The t-test investigates the following hypothesis:

For null hypothesis (H_o): $\mu_1 = \mu_2$

For alternative hypothesis (H_a): $\mu_1 \neq \mu_2$

t-statistic is given by the equation (4.3)

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{4.3}$$

where $\bar{y}_1$ and $\bar{y}_2$ are the means of the two samples, s_1 and s_2 are the standard deviations of the two samples, and n_1 and n_2 are sample sizes.

3.5.2 Wilcoxon test

Wilcoxon test is a non-parametric (Mann and Whitney 1947, Wilcoxon 1945) equivalent of parametric t-test. The non-parametric test assumes that samples are randomly taken from population with symmetric frequency distribution and does not require that data to follow normal distribution. The test investigates hypothesis on median and can be applied on single sample or two samples (paired or unpaired samples). In single sample case, the test investigates whether the median of sample is different from the hypothesised median of the population. In two samples case, Wilcoxon test investigates median of one sample is different from the second one. Two common non-parametric tests are: Wilcoxon signed-rank test for paired data and the Mann-Whitney U test (also known as Mann-Whitney-Wilcoxon test, the Wilcoxon T test, the Wilcoxon two-sample test, or the Wilcoxon W test) for unpaired data. These tests are based on ranking of the data and looking at the ranks rather than the actual values of the observations.

3.5.3 Analysis of variance

Analysis variance (ANOVA) is used to compare the means of two or more groups using F-statistic under the assumption that sampled population are normally distributed (Snedecor and Cochran 1989). One-way ANOVA allows determining whether one given factor (factor is an independent variable whose values are controlled and varied, for example, in experiments) has significant effect in mean values of any groups in the data.

F-test statistic computes ratio of two sources of variability as below:

F = between group variability / within group variability

ANOVA tests the following hypothesis:

Null hypothesis (H_o): $\mu_1 = \mu_2 = \mu_3 ... = \mu_k$

Alternative hypothesis (H_a): Means of all groups are not equal.

The significant *p* value means that there is at least one group whose mean is different from the rest of groups. One-way ANOVA, however, does not provide information on which group is different from the rest. Post-hoc tests are needed to find which specific group(s) is different from the rest.

3.5.4 Multiple hypothesis testing

The probability of making Type I errors increases rapidly along with the number of hypotheses tested simultaneously. This is called multiple hypothesis testing problem. This problem has become routine in *omics* studies (Farcomeni 2008) where large numbers of statistical tests are performed in the same data set independently on a number of variables. This is the condition where one needs to account for the multiple tests performed. For instance, when employing a t-test for comparison of means across two groups at 5% significant level, the test is willing to accept 5% error i.e., if 100 hypothesis tests are performed, it is expected to see five significantly different variables by chance alone even if there are no actual differences. Many solutions (e.g., Anisimova and Yang 2007) are suggested to account for multiple hypotheses testing including Boneferroni and false discovery rate approaches. Boneferroni correction (Miller 1981), which controls family wise error rate, is the simplest and more conservative correction to account for the multiple hypothesis testing. This correction obtains acceptable significant level by taking into account number of hypothesis tests performed. This is achieved by dividing the p-value of the test by the number of tests performed. Other notable multiple hypothesis testing correction is false discovery rate (FDR) (Benjamini and Hochberg 1995) which is the expected proportion of Type I errors among the rejected hypotheses. It is less conservative approach as compared to family wise error rate correction, which is the probability of making at least one Type 1 error over all hypothesis tests.

3.5.5 Correlations

Correlation describes the degree of relationship between two variables (X and Y) and is measured using correlation coefficient. The value of correlation coefficient varies from 0 (no relationship between X and Y) to 1 (perfect linear relation ship) or -1 (perfect negative linear relationship). A positive value for the correlation implies a positive association (large values of X tend to be associated with large values of Y and small values of X tend to be associated with small values of Y). A negative value for the correlation implies a negative or inverse association (large values of X tend to be associated with small values of Y and vice versa).

The most common measure of correlation is Pearson correlation (Pearson 1896) which is computed using equation (4.4)

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \tag{4.4}$$

where r_{xy} is correlation coefficient between X, Y variables, x_i, y_i are series of measurements on X and Y respectively, S_x and S_Y are standard deviations of X and Y respectively and $\bar{x}$ and $\bar{y}$ are sample means of X and Y respectively.

The non-paramteric version of measuring correlations is Spearman's rank correlation (Spearman 1904) which is computed as below:

$$\rho = 1 - \frac{6\sum d_i^{\,2}}{n(n^2-1)} \tag{4.5}$$

where ρ is Spearman's rank correlation coefficient, n is the number of values in each data set and d_i is the difference between the ranks of corresponding values X_i and Y_i.

4. RESULTS AND DISCUSSION

This chapter covers main results related to informatics framework for non-targeted screening of lipids, computational aspects of functional class label prediction for unidentified lipids, lipid pathways instantiation strategy and exploratory analysis of lipidomics data. More detailed information can be found in the original publications **I–VI**.

4.1 Lipid identification

One of the main challenges in lipidomics is to characterize the complete lipid inventory present in a given biological matrix. This challenge motivated the present work to compute spectral libraries for the screening of potential lipid species. This was important especially as there were no publicly available inter-laboratory transferable lipid libraries for the LC/MS platforms.

4.1.1 Database content and basic search

Lipid database (LipidDB) was constructed computationally using SMILES approach (Publication **I**). The LipidDB comprised main classes of lipids such as glycerophospholipids, sphingolipids, glycerolipids, and sterol esters. Specific contents of each lipid class are as shown in the Table 4.1. The enormous structural diversity found in these classes of lipids is due to the differences in length and degree of unsaturation in alkyl chains. Structural rules of specific class (i.e., glycerophospholipids) typically follow a common template which allowed incorporating the structural diversity computationally. The computational framework was based on the construction of “seed” fatty acids most likely to occur in living systems. Each lipid entry was assigned a scoring value based on the seed fatty acid composition to facilitate the searches of experimental results against LipidDB. The scoring value aided when search results were associated

with multiple hits due to isobaric and/or isomeric matches in mass as well as limitations in the analytical approach. Such a heuristic scoring scheme can be modified and the scheme may be different for different cell or tissue types.

Table 4.1 List of different lipid classes and their specific contents in the LipidDB.

Class	**Description**
Fatty Acyls	Alcohols/aldehydes/carboxylic acids and CoAs
Glycerolipids	Mono acyl/alkyl glycerols Diacyl/alkyl glycerols Triacylglycerols
Glycerophospholipids	Glycerophosphocholines, glycerophosphoethanolamines, glycerophosphoserines, glycerophosphates, glyceropyrophosphates and glycerophosphoglycerols.
Sphingolipids	Sphingoid bases, various ceramides including ceramide phosphoinositols, ceramide phosphocholines, ceramide phosphoethanolamines, N-acylsphingosines, N-acylsphinganines, ceramide 1-phosphates and sulfatides.
Sterols	Cholesteryl esters
Plasmalogens (glycerophospholipids)	Phospholipids with vinyl ether bonds at sn-1 position

LipidDB was stored in a native XML database implemented in Tamino XML Server (Software AG). Each lipid entry in LipidDB was described by an internal identifier, scoring information, class, canonical SMILES, molecular formula, molecular weight and isotopic distribution. XML schema for LipidDB is shown in the Figure 4.1. This database was extensively used for the extraction of crucial information using basic search interface as shown in Figure 4.2. The basic search allows queries on lipids species based on their molecular masses, adduct information, lipid classes, fatty acid chain positions, head group information and number of bonds.

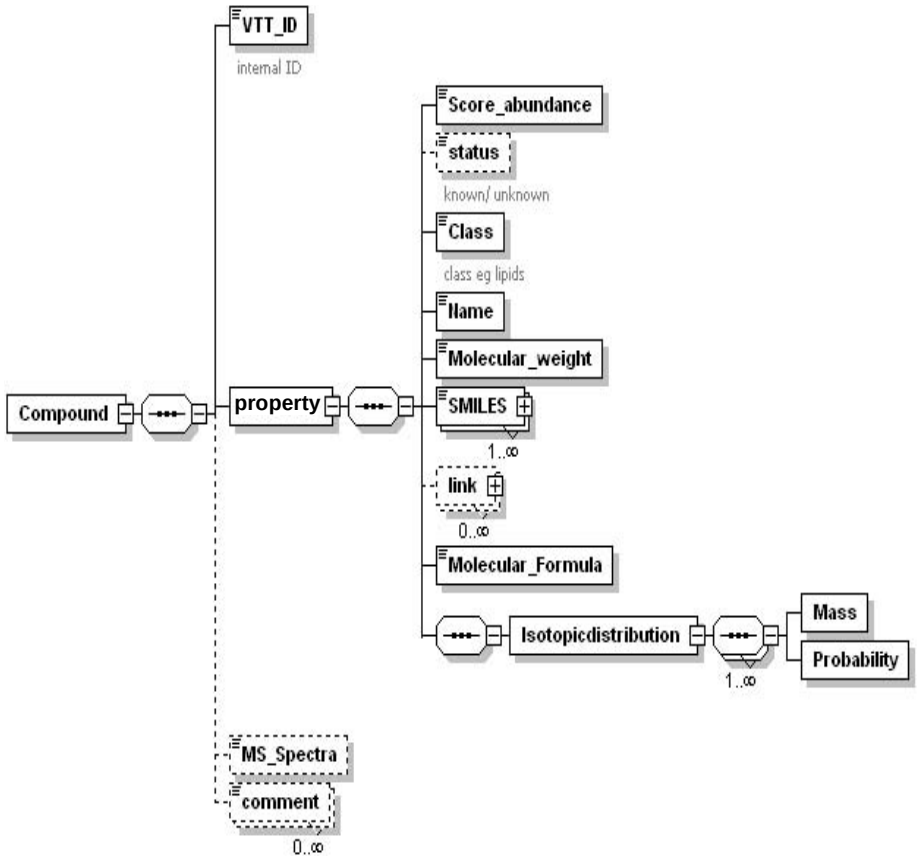


Figure 4.1. XML database schema for LipidDB stored in the native Tamino XML database.

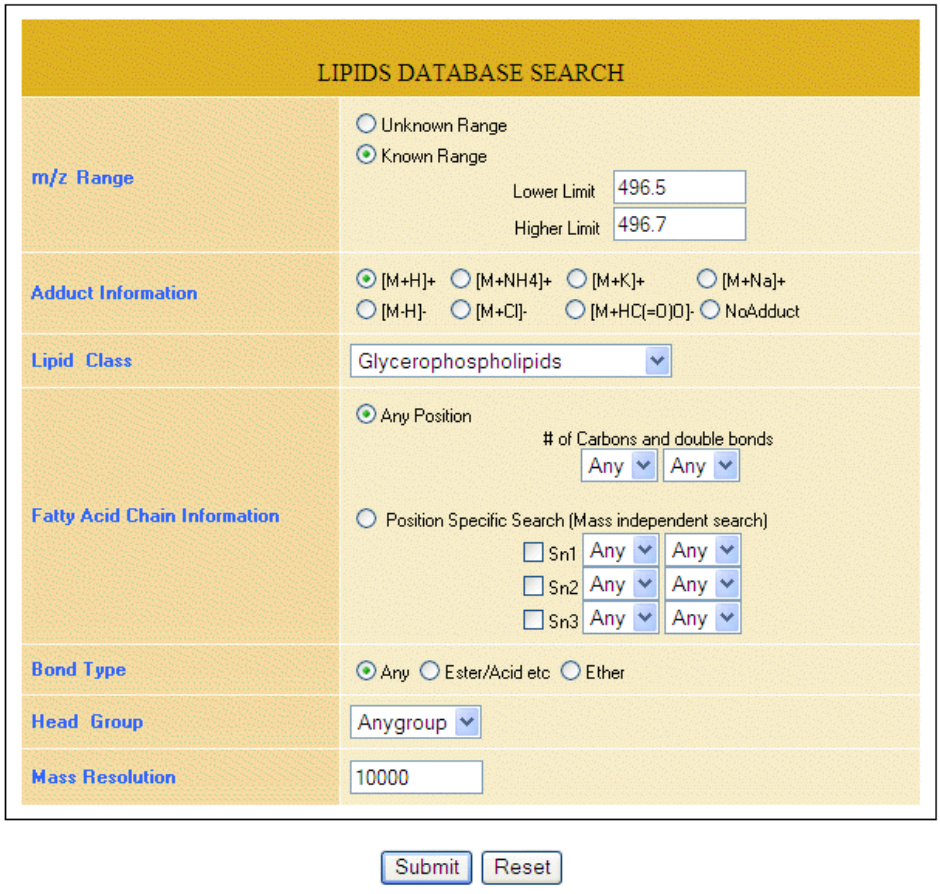


Figure 4.2. Basic search interface for the extraction of lipid information from the underlying LipidDB.

4.1.2 Customisation of LipidDB for UPLC/MS platform

Customisation of experimental information such as retention time, adducts and MS/MS fragmentation in *in silico* database is useful for the screening of potential lipids (Figure 4.3).

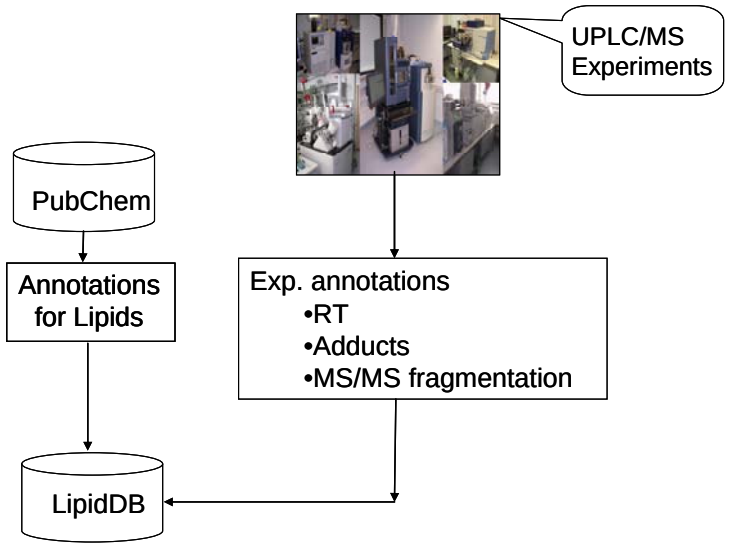


Figure 4.3. LipidDB was updated with external annotations from public databases as well as experimental information from UPLC/MS platform.

Annotations from external databases: Publicly available databases such as PubChem (http://pubchem.ncbi.nlm.nih.gov/) provide valuable information on small molecules including lipids. LipidDB was annotated with physical and chemical properties as well as external IDs of different databases available in the PubChem database.

Addition of retention time ranges and adducts information: In UPLC/MS-based global lipidomics screening, retention time range information is useful parameter and can serve as a coarse filter to avoid false positive hits. Detailed tandem mass spectrometry (MS/MS)-based characterization of all peaks revealed information about retention time ranges for several classes of lipids. The information was obtained from the two dimensional m/z and retention time plot (Figure 5, Publication **II**) generated using MZmine software version 0.60. Characterization of class specific regions and their adduct information was based on MS/MS spectra.

Addition of MS/MS fragmental information: In order to facilitate the identification of lipid species, the main fragmental peaks of acyl glycerols, phospholipids, cholesteryl esters and sphingolipids were included in *in silico* LipidDB. The computational library had greatly facilitated lipid identifications

while comparing with experimental MS/MS spectra. The main fragmentation information of different classes of lipids is summarised as below:

Glycerophosphocholine (PC): In the positive ion mode, PC molecular species form either protonated [M+H]+ or sodiated [M+Na]+ adduct ions as well as dominant peak at m/z 184, representing choline head group. Sodiated PC molecular species produce characteristic fragments at m/z [M+Na-59]+, [M+Na-205]+ and [M+Na-183]+. To facilitate the assignment of fatty acid moieties at sn-1 and sn-2 positions, negative ion mode ESI-MS/MS analysis was performed. In general, negative ion spectra of all phospholipids classes yield four series of ions (Pulfer and Murphy 2003) corresponding to (1) loss of fatty acyl substituents as free fatty acids (2) loss of fatty acyl substituents as ketenes (3) fatty acyl carboxylate anions and (4) head group specific ions. In the negative ion mode, PC molecular species yield ions characteristic for formate adduct [M+HCOO]- and demethylated [M-CH3]- species as well as ions characteristic of the fatty-acyl group esterified at the sn-1 and sn-2 positions. Product ion spectra of plasmanyl (alkyl ether linkage at sn-1 position) / plasmenyl (vinyl ether linkage at sn-1 position) molecular species contains information predominantly related to fatty acyl chain at sn-2 position as well as ion reflecting the loss of sn-2 fatty acyl chain and hence are distinguished from ester-linked phosphatidyl species (Khaselev and Murphy 2000, Zemski Berry and Murphy 2004).

Glycerophosphoethanolamine (PE): Phosphatidylethanolamine, being zwitterionic, can be detected both in positive and negative ion mode mass spectra. Characterization of molecular species as their protonated species [M+H]+ and subsequent yielding of major fragment at m/z [M+H -141]+ (due to loss of head group) is used for the identification of PE molecular species in the positive ion mode. Ethanolamine plasmalogens are detected based on two fragment ions characteristic of sn-1 and sn-2 positions (Khaselev and Murphy 2000, Zemski Berry and Murphy 2004). In the negative ion mode, PE molecular species form deprotonated ([M-H]-) ion which undergoes cleavage of fatty acyls substituents mainly as ketenes. Head groups specific ions are observed at m/z 140 (phosphoethanolamine ion) and 196 (i.e., loss of fatty acyl groups in PE).

Glycerophosphatidylserine (PS): In the positive ion mode, PS is detected as protonated ion. Structural characterization of PS species is mainly done using negative ion mode ESI-MS/MS. In this mode, PS species form [M-H]- ions and [M-H-87]- ions, arising from the loss of serine group upon fragmentation as well as ions corresponding to loss of fatty acyl substituents as ketenes.

Glycerophosphoglycerol (PG): These are less abundant ions of phospholipids. In the negative ion mode, these molecular species yield [M-H]- ions as well as characteristic peaks of lysophospholipids-like fragments due to loss of fatty acyl ketenes/acids and carboxylate anions. Head group specific fragments are detected at m/z 227 and 171.

Glycerophosphoric Acid (PA): PA is the simplest phospholipid and preferentially studied in the negative ion mode in which PA yields deprotonated ion ([M-H]-). Like in other phospholipids, PA forms ions corresponding to neutral loss of acids, neutral loss as ketenes and carboxylate anions. Head group specific ion is detected at m/z 153, a characteristic ion arising from loss of fatty acyls groups from PA species.

Glycerophosphoinositol (PI): Negative mode ESI-MS yields [M-H]- ions of PI. The major fragmentation pathways involves neutral loss of fatty acid, neutral loss as ketenes and loss of the inositol head group (m/z 162, inositol − H2O). A prominent characteristic ion at m/z 241 represents a dehydrated product of inositol phosphate.

Triacylglycerol (TG): ESI-MS of TG species yields ammonium adduct ions which are fragmented in MS/MS to diacylglycerol ([DG]+) like fragments and are similar to those of [DG]+ species in phospholipids due to loss of head groups in phospholipids. These [DG]+ species are informative in identifying TG species. However, in the analysis of mass spectra with co-eluting TG species, it is difficult to assign [DG]+ fragments to its parent TG molecular species correctly.

Cholesteryl Ester (ChoE): ESI-MS platform is not well suited for the analysis of free cholesterol. ChoEs, however, form ammonium adducts in the positive ion mode and generate a fragment ion at m/z 369 upon collision-induced fragmentation.

Sphingolipids: In the positive mode, ESI-MS analysis of sphingomyelin (SM) yields a characteristic protonated phosphocholine peak at m/z 184. PC and SM species are distinguished based on their characteristic m/z value (PC species occur at even at m/z and SM species at odd m/z). Similar to PC, sphingomyelin yields either [M+H]+ or [M+Na]+ ions in the positive mode, while in negative ion mode are [M-CH3]- and [M+ HCOO]- ions.

In the positive ion mode, ceramides form unstable protonated molecular species which undergo dehydration to form [M+H-H_2O]+ ion. Molecular ions in negative ionisation conditions are very informative in identifying the fatty acyl and long chain base substituents of ceramide. Ceramide species yields [M-H]- and [M-H-30]$^-$ (due to loss of HCHO group) ions in negative ionisation conditions. While positive mode analysis of long chain bases such as

sphingosine (d18:1), sphinganine (d18:0) and 4-D-hydroxysphinganine (t18:0) undergo dehydration to form fragments at m/z 282/264, 284/266 and 300/282, negative ion mode analysis of sphingosine and phytosphingosine moieties are characterised by the fragments at m/z 237/263 and 225/255/267 respectively (Merrill et al. 2005, Myoung Hee Lee and Jong 2003).

4.1.3 Building of tissue-specific lipid libraries

Comprehensive mass spectrometry studies allow building of MS/MS fragment libraries for different biological tissues. These libraries are useful for intra-laboratory use in rapidly assigning lipid species coming from new experiments under similar conditions. Tissue-specific libraries were built using the fragmentation characteristics of lipids described in the section 4.1.2. Such MS/MS spectral libraries have limitations due to laboratory-dependent ion source conditions. Moreover, ESI usually produce little structural information and single set of conditions are not applicable for broad spectrum of lipids, or metabolites in general.

Lipoprotein lipidomics in the context of insulin resistance and abdominal obesity

Background: In order to investigate the relation between serum lipid and lipoprotein abnormalities with insulin resistance, sixteen non-diabetic subjects between 18 and 60 years of age were recruited based on a healthy clinical background and modest alcohol consumption. Relevant clinical parameters of all subjects were measured using standard protocols as described in the original publication (Publication **VI**). The lipoprotein fractions such as VLDL, intermediate density lipoprotein (IDL), low density lipoprotein (LDL) and high density lipoprotein (HDL) were separated by sequential flotation in an ultracentrifuge (Taskinen et al. 1988).

In contrast to the traditional measurement of total protein, phospholipid, cholesteryl esters, and TG content in a given biological sample (Vance and Vance 2008), modern MS-based techniques allow analysis at the molecular species level. The MS-based methods have become mainstay of lipidomic research mainly with two strategies: global and targeted approaches. Global (or non-targeted) approaches are directed towards identification and quantification of several hundreds of lipids in a high-throughput basis. In this direction, multiple shotgun-based MS approaches (Ejsing et al. 2009, Han & Gross 2005b) have been developed for wider coverage of different lipid classes. Our recent analysis using UPLC/MS-based platform allowed the analysis of multiple

abundant lipid classes (e.g., triacylglycerols, cholesterol esters, sphingomyelins, phosphatidylcholines) as well as bioactive lipid species (e.g., ceramides, plasmalogens, and lysophosphatidylcholines) simultaneously (Pietiläinen et al. 2007). This global approach does not assume any prior knowledge on type of lipids to be screened and thus providing greater possibility for discovering new classes of lipids. The quantification of lipids in the presented methodology requires further optimization given that lipids cover wide concentration range of compounds. The targeted approaches have been developed for screening of one or few classes of lipids and are more quantitative. Like any other lipidomics methods, the UPLC/MS-based method has limitations in detecting the position and configuration of double bonds within the fatty acid moieties of lipid species. Recent developments on ozone-induced dissociation (OzID) of double bonds (Thomas et al. 2007) can offer a potential solution.

Lipid characterization: Non-targeted profiling was performed to determine the individual species in each lipoprotein fraction. Tandem mass spectrometry was performed both in positive and negative ion modes. Negative mode analysis was mainly utilized to determine fatty acid composition of phospholipids. This extensive study, typically involved examining each individual product ion spectra and compiling spectral information, allowed building of spectral libraries for lipoprotein fractions.

The comprehensive profiling allowed us to detect the compositional details of bioactive lipid species in different lipoproteins. Ceramides were found only in VLDL and LDL, whereas ethanolamine plasmalogens (PE(p)) were found only in LDL and HDL2. Lysophosphatidylcholines (lysoPC) and ether linked phosphatidylcholines (PC(e)) were present in all lipoproteins with the greatest abundance in HDL2, HDL3, and LDL (Figure 2, Publication **VI**). Identification of individual TGs in major lipoprotein particles allowed us to elucidate how changes in different TGs and fatty acids related to features of insulin resistance and abdominal obesity (Figure 1, Publication **VI**).

As a summary, the SMILES-based approach allowed the construction of lipid database for identification of lipids in mass spectrometry analysis. The direct application of databases as described in section 2.4 for mass spectrometry is limited due to the nature of LC/MS-based analysis. The LC/MS-based spectral libraries are very much dependent on the type of scanning mode and instrumental settings and hence it is very unlikely any single database can act as a standard reference for all types of lipidomic analysis. This is also partly due to

large diversity of lipids across different organisms, tissues, and cell types. Here, the developed database was customized to global screening methods by addition of RT information, adducts, mass fragments and other annotations from external databases. The database was extensively used in the construction of tissue specific spectral libraries.

4.2 Functional class label prediction of unidentified peaks

Mass spectrometry-based metabolomics experiments often results in unidentified peaks which can hamper in the interpretation of results. The problem is even more challenging in the non-targeted metabolome screening experiments since modern MS instruments have ability to detect several hundreds of peaks in a given sample. Even when the identification is relatively easier for lipids if MS/MS spectra are obtained, non-targeted UPLC/MS approach often results in unidentified peaks due to analytical limitations resulting from small peaks, co-fragmentation, ambiguous spectra, as well as complex spectra probably coming from modified and uncommon lipids. In fact, data analysis (univariate or multivariate analysis) often lead to interesting peaks which are often unidentified and may serve as potential biomarkers. Strategies are therefore needed to interpret the data when exact identifications are not available. The challenge was addressed computationally by predicting functional class labels for unidentified peaks (Publication **II**). This methodology serves as a helpful intermediate step in data analysis as well as a guide towards the further steps to identify the compounds.

Computational methodology was demonstrated using the lipidomics data from our earlier twin pair study (Pietiläinen et al. 2007). The lipidomic data were preprocessed using an MZmine software version 0.60 (Katajamaa and Orešič, 2005, Katajamaa et al., 2006). Based on identifications, lipids were assigned to one of the following classes: glycerophosphocholines (PC) glycerophosphoethanolamines (PE), sphingomyelins (SM) and triacylglycerols (TG). Cross validation strategy was employed to assess the generalisation performance of the classifiers. Here, computational work involved single cross validation method for models with no meta-parameter estimation (i.e., random model and Naive Bayes) and double cross validation for models requiring metaparameter estimation (i.e., k-NN, SVM and PLS-DA). Details of single and double cross validation methods are available in original publication (Publication **II**). Main results of employed supervised classifiers (PLS/DA, SVM, Naive Bayes and k-NN) are summarised below.

4.2.1 PLS/DA

PLS-DA is a latent variable-based supervised classifier and was investigated to evaluate the performance in predicting the functional labels for unlabeled peaks. Repeated double cross validation setting was employed to estimate the optimal number of latent variables, which was found to be mainly between 5 and 20 components. The optimum number of components from inner loop of double cross validation was used to evaluate the performance of classifier on corresponding test sets in the outer loop. The errors made in predicting test set labels were computed and overall accuracy of the PLS-DA classifier was found to be 63.3%. In all, the classifier performed slightly better than both Naive Bayes and random models and underperformed as compared to k-NN and SVM models (Table 1, Publication **II**)

4.2.2 SVM

SVM classifier was used to solve multiclass-classification problem using 'one-against-one' approach which trains L(L-1)/2 (L = number of levels) binary classifiers. The appropriate class was found by the majority voting scheme. Linear kernel function was utilised and the performance of classifier was evaluated using four-fold cross validation study with double cross validation being repeated for 25 times. Double cross validation was used to select an optimal regularisation parameter (C). The parameter corresponding to minimum cross validation error was varied mostly between 100 and 1000. The cross validation accuracy of SVM classifier was found to be 92.83 %. Prediction accuracies of SVM model on lipid classes were found to be better than those predicted by k-NN and Naive Bayes models (Table 1, Publication **II**).

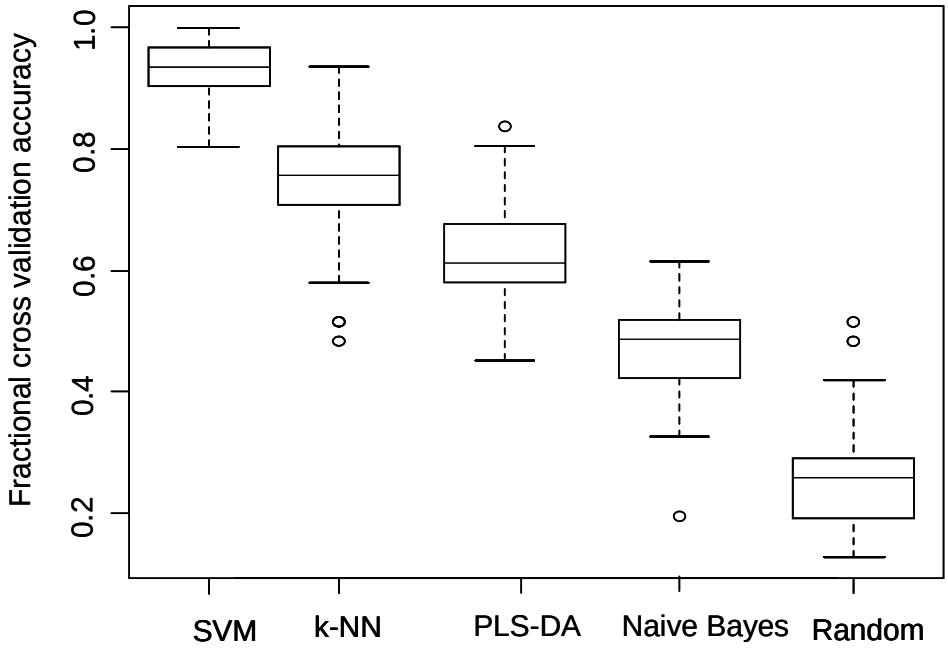


Figure 4.4. Comparative cross validation performances of SVM, k-NN, PLS-DA, Naive Bayes and random classifiers.

4.2.3 Naive Bayes

Naive Bayes classifiers are probabilistic version of classifiers. It was also trained using four-fold cross validation repeated 25 times. Cross validation accuracy in class label prediction on test set as well as individual classes are shown in Table 1 of Publication **II**. The performance based on cross validation accuracy of Naive Bayes model was found to be poor as compared to k-NN and SVM models.

4.2.4 k-NN

k-NN classifier requires estimation of parameter k to decide appropriate class label for a given test sample. The value of k represents the number of nearest training samples in the feature space considered when deciding class label for the test sample. Repeated four-fold cross validation procedure was employed to select optimal k and performance evaluation of the k-NN classifier. The average cross validation accuracy of k-NN classifier for the whole test set as well as for

each lipid class was summarized in Table 1 of Publication **II**. Cross validation accuracy of k-NN model was 75.97 %. k-NN classifier outperformed random assignment model both in terms of class-specific as well as over all cross validation performance. Naive Bayes model showed poor cross validation performance as compared to k-NN model on complete test set as well as TG and SM classes. Better prediction accuracy of lipid classes (TG, PC and SM) demonstrated predictive ability of k-NN classifier in lipid data.

Here we attempted to make use of unidentified peaks in the lipidomic analysis by predicting the functional labels using supervised classifiers. The k-NN and SVM classifiers outperformed Naive Bayes and PLS-DA classifiers. More robust prediction could be achieved by utilizing consensus predictions from both SVM and k-NN classifiers. The poor performance of Naive Bayes classifiers could be attributed to class-specific co-regulations of lipids. This computational framework complements the existing identification methodologies with predictions of class labels to facilitate exploratory analysis. These kinds of appproaches are more useful in the global screening approaches in the metabolomics where the identification of metabolites is a bottleneck.

4.3 Reconstruction of lipid pathways

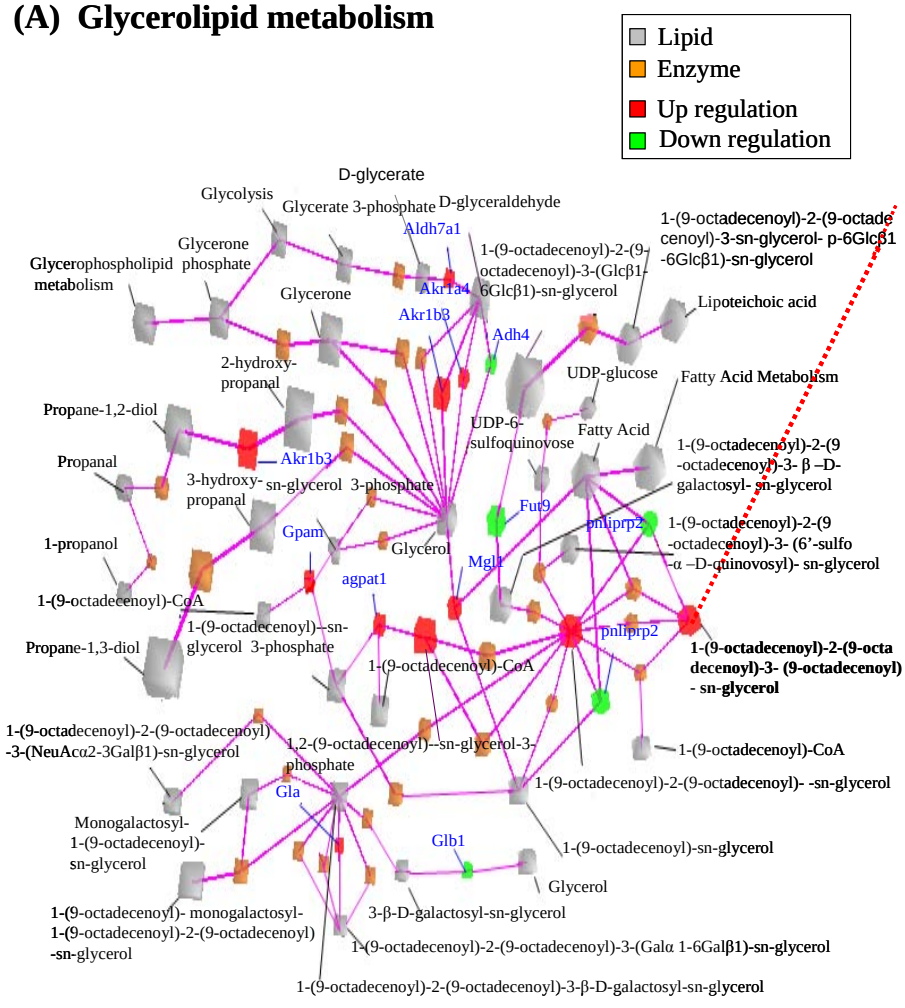
Biochemical pathways are rich sources of information and may help to gain mechanistic links behind underlying phenotype. Properly organized and curated databases are required to view or construct biochemical pathways reliably. Databases such as KEGG PATHWAY database provide information on available lipid pathways besides other biochemical pathways. The database also provides generic, organism-independent biochemical pathways that serve as reference pathways for constructing organism-specific pathways. The available databases therefore serve as a good starting point for pathway analysis.

4.3.1 Pathway instantiation

Biochemical research is empowered with modern analytical techniques which can provide plenty of detailed molecular information. As a result, available pathway databases need to accommodate these changes to the resolution of available information. This problem is obvious particularly in the case of lipids. The essential building blocks for the molecular pathway instantiation were explained in the original publication (Publication **I**).

Each node in Figure 4.5 is either a lipid metabolite or enzyme or other interconnecting metabolic pathway. Grey color represents metabolites/other metabolic pathways and brown represents enzymes. Up- and down-regulation of corresponding nodes are denoted by red and green colors, respectively. Enzyme names are shown only if they are differentially regulated (~ 1.5 fold change). Glycerolipid pathway instantiation was demonstrated with TG(18:1/18:1/18:1) (Figure 4.5A) lipid species where experimental measurements could be mapped directly on pathway unlike in generic pathways and thus bridging the gap between MS data and existing lipid pathways. From the sphingolipid pathway map (Figure 4.5B), two enzymes linked to the ceramide via metabolic reactions, one is SGPP1 (Sphingosine-1-phosphate phosphatase 1, UniProt ID Q9JI99), the other GALC (galactosylceramidase, UniProt ID P54818) were upregulated in ob/ob. SGPP1 is involved in *de novo* ceramide synthesis, while GALC hydrolyses galactosylceramide to form ceramide. Interestingly, sphingomyelin SM(d18:1/18:0) a precursor of ceramide via the sphingomyelinase reaction is downregulated, while the sphingomyelinase level is maintained. Therefore, these results indicate that both glycolipids and free fatty acids may act as a source of the elevated ceramides in the ob/ob fatty liver.

(A) Glycerolipid metabolism



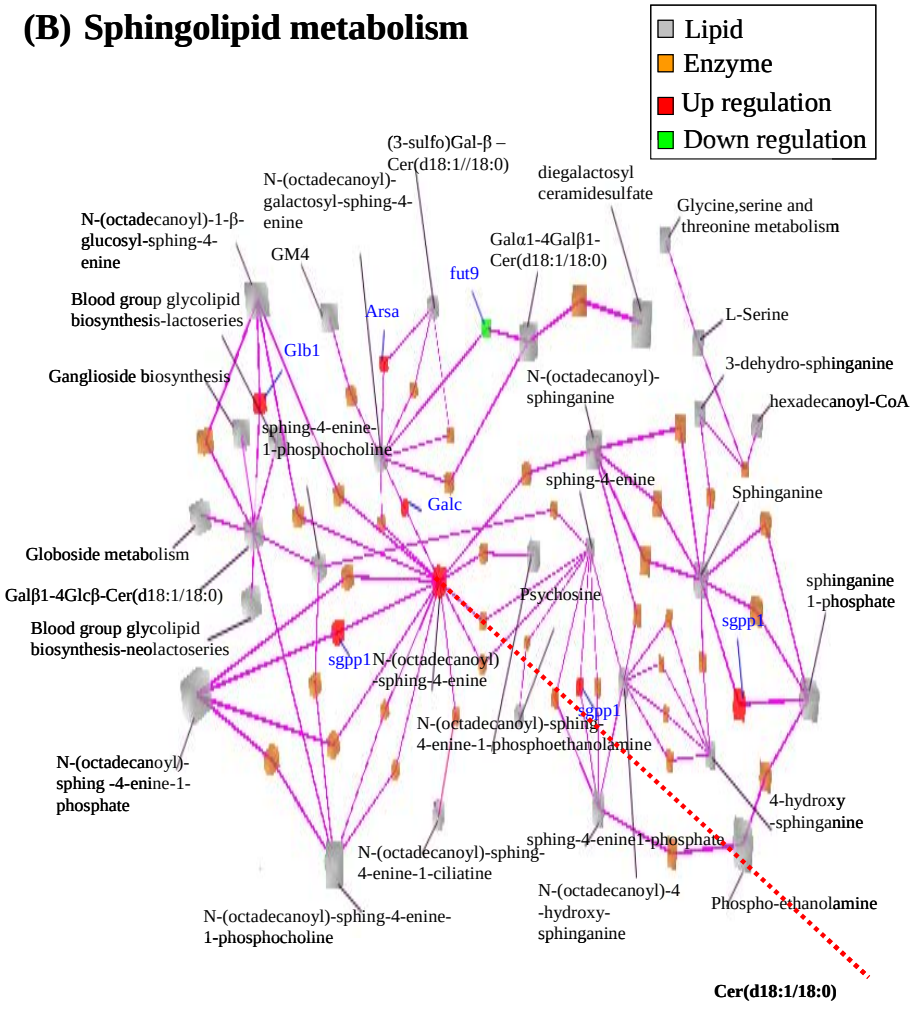


Figure 4.5A, B. Instantiated pathways for Cer(d18:1/18:0) and TG(18:1/18:1/18:1). A) Instantiation of TG(18:1/18:1/18:1) as part of glycerolipid metabolism. (B) Instantiation of Cer(d18:1/18:0) as part of sphingolipid metabolism.

4.3.2 Tissue- and context -specificity in lipid pathways

In order to facilitate any modifications in the existing pathways, the megNET visualisation tool is extended to construct pathways for a given list of EC numbers and lipid data. The developed frame work is now being targeted towards building context-sensitive and tissue-specific pathways. As a part of new developments, a new database was compiled with all available lipid-related

proteins with their pathway information such as KEGG pathways and context information such as gene ontology terms. Knowledge about the tissue-specific enzymes was incorporated at transcriptomic level. The presence or absence level information of mRNA in tissues is expected to facilitate the development of tissue-specific pathways.

4.4 Lipid profiling applications and data analysis

High-throughput molecular profiling technologies provide an opportunity to measure lipids on an unprecedented scale. The large amount of data presents a major challenge for statistical methods to handle and assemble proper knowledge for biomarkers discovery efforts. These analyses typically start with unsupervised methods where the main aim is to get accurate knowledge on samples whether they really cluster or provide trends in the data. These techniques also serve as good visualisation tools in data analysis. Unsupervised methods such as principal component analysis (PCA) are useful to capture the trends mainly when there are correlated variables in the data (e.g., lipid data). Such PCA-assisted analysis of projecting samples into lower dimensional space from high dimensional space not only allow one to confirm the expected patterns in the data based on the group membership of samples but also facilitates in detecting outliers in samples. Supervised methods such as PLS/DA can also be employed to describe complex data with few latent components as well as a obtain set of most distinguishing variables (lipids) among the groups of samples.

4.4.1 Lipidomic profiling of multiple tissues of the POKO mice

Increased obesity is the one of the risk factors for type 2 diabetes. The relationship of how obesity causes the diabetes still remains unknown. It is hypothesised that when adipose tissue reaches its full capacity, the excess fat spills over to other metabolically active organs such as liver, pancreas and skeletal muscle. This condition leads to insulin resistance and diabetes (Gray and Vidal-Puig 2007). Earlier studies indicate that peroxisome proliferator activated receptor gamma (PPARγ) plays a key role in adipogenesis and insulin sensitivity (Koutnikova et al. 2003, Rosen et al. 1999, Spiegelman 1998). However, the importance of PPARγ2, which is nutritionally regulated isoform of PPARγ, is still not clear.

In order to investigate the physiological importance of PPARγ2 under positive energy balance conditions in ob/ob mice (PPARγ2þ/þ Lepob/Lepob) (Publication **V**), lipidomic analyses were performed on relevant tissues. Profiling of adipose tissue, pancreatic islets, liver, and skeletal muscle samples revealed distinct differences in four genotypes: WT (PPARγ2þ/þ Lepþ/Lepþ), PPARγ2 KO (PPARγ2_/_ Lepþ/Lepþ), ob/ob (PPARγ2þ/þ Lepob/Lepob), and POKO (PPARγ2_/_ Lepob/Lepob). The adipose tissue from POKO mice was characterised by decreased triacylglycerols (TGs) and increased diacylglycerols (DGs). These changes were associated with increased levels of ceramides. Lipid profiling revealed decreased TG and DG levels and increased ceramides (Cer) levels in POKO islets, indicating the possible role of PPARγ2 in promoting TGs levels to increase the lipid-buffering capacity of β-cells thereby preventing lipotoxicity. Liver and skeletal muscle lipidomics revealed decreased TGs and increased formation of bioactive lipid species such as ceramides and lysophosphatidylcholines in POKO mice compared to ob/ob mice. In all, lipidomic profiling of four tissues showed similar pattern of changes.

The study on 16-week-old mice revealed increased levels of ceramide in POKO islets as compared to *ob/ob* islets. The study was later extended to investigate whether perturbed lipid metabolism is already present in islets at 4–5 weeks of age (Publication **IV**). There were no statistically significant changes in lipid composition among the four genotypes of mice (Figure 4.6) unlike in the 16-week-old mice. The statistical significance was based on one way analysis of variance (1-way ANOVA) and associated p value was adjusted for multiple hypothesis testing. Analysis in other metabolically active tissues such as serum, liver, adipose tissue and muscle from WT, PPARγ2 KO, *ob/ob* and POKO mice, however, showed significant lipid compositional changes in 4-week-old mice**.** In serum, POKO mice had higher levels of TGs and high levels of short- and medium-chain PC species as compared to mice with other genotypes. Interestingly, the levels of long-chain TGs were lower in ob*/ob* mice than in WT and PPARγ2 KO mice. Lipidomic characterization of liver tissue revealed increased TG levels (mainly short- and medium chain TGs) at 4 weeks of age in both POKO and *ob/ob* mice as compared to WT mice. Ob/ob and POKO mice livers had increased levels of medium chain PCs when compared with WT and PPARγ2 KO mice. Unsaturated long-chain TGs were enriched in POKO and *ob/ob* livers. Lipidomic profiling in adipose tissue revealed similar levels of short-, medium- and long-chain TGs in POKO and *ob/ob* mice. Polyunsaturated long-chain TGs were enriched in POKO and *ob/ob* mice as compared to WT and

PPARγ2 KO mice. Lipidomic data from skeletal muscle showed that 4-week-old ob/ob mice had increased short- and medium-chain TGs as compared to other genotypes. Muscle from POKO and PPARγ2 KO mice contained more long-chain TGs than WT and *ob/ob* muscles. Interestingly, the levels of ceramides and lysoPCs increased similarly in both *ob/ob* and POKO mice**.** Taken together, lipdomic analysis of relevant tissues from 4-week-old mice suggested abnormal accumulation of TGs and the resulting lipotoxicity may contribute to the severity of the metabolic syndrome in 16-week-old POKO mice**.** These studies and more specifically, the differences observed between the liver, serum and muscle, may also indicate the possibility of a hierarchical order of organs with respect to fat deposition and lipid-induced toxicity.

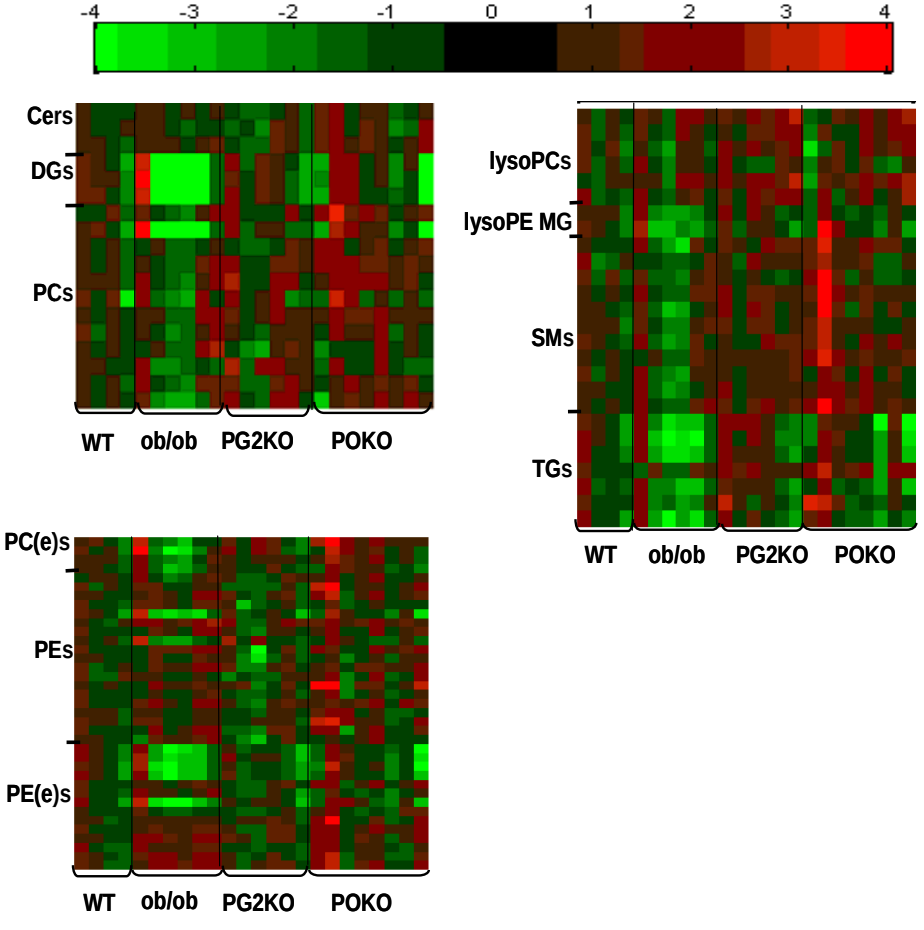


Figure 4.6. Lipidomic profiling of 4-wk-old mice islets from wild type, PPARγ2 KO, ob/ob and POKO mice (n = 5–8). No significant lipids with ANOVA p-values < 0.05 were found.

We further investigated the 4-wk-old lipidomics data using chemometric approaches to find which specific lipid variables are responsible for the separation of the four genotypes. Serum lipid profiles from four genotypes were used to demonstrate the usefulness of exploratory analysis. PCA analysis was performed to detect outlier samples and check whether samples of same genotypes are clustered together. X-block was autoscaled prior to PCA analysis. Separation of four genotypes was not very clear (Figure 4.7A) and more fine clusters were obtained using PLS/DA analysis. Outliers detected in PCA were removed prior to PLS/DA. The data were preprocessed by autoscaling X-block data and mean-centering Y-block data. Cross validation (contiguous block cross validation method) and Q^2 scores were used to optimise the PLS/DA model.

The variable importance in the projection (VIP) values (Wold et al. 1987) were computed to identify most important lipid species contributing to separation of four genotypes in the PLS/DA model. Top scoring VIP lipid variables responsible for separation each group were found. For instance, VIP plot of serum lipidomics for wild type is shown in Figure 4.8. Top VIP scoring lipids from the plot are SM(d18:1/22:0), SM(d18:1/24:1), TG(58:9), PE(36:2) and PE(38:6e). Similar VIP analysis was also performed on other genotypes to find their respective variables responsible for the separation among the genotypes.

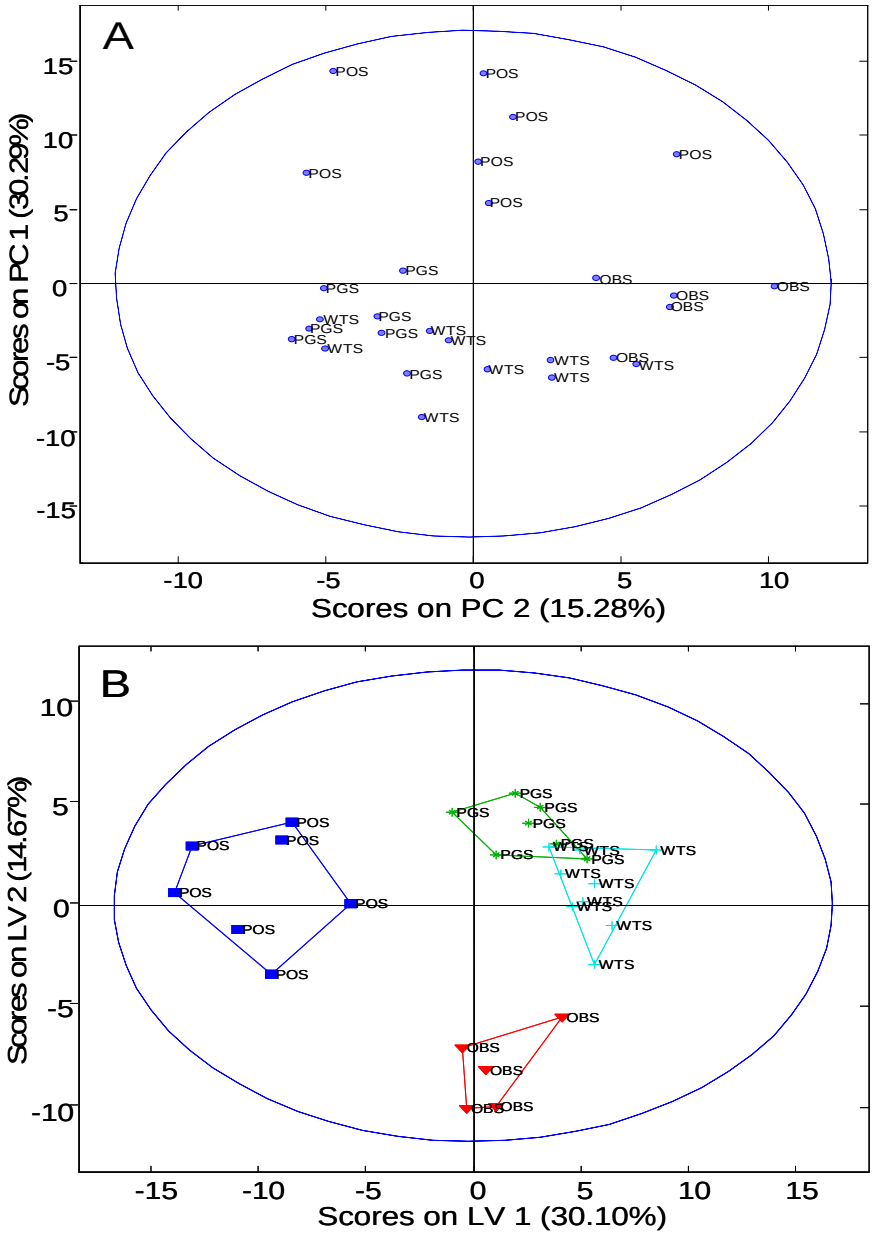


Figure 4.7A, B. Exploratory analysis of serum lipidomic data in the ob/ob mice model for four genotypes. Legends WTS, PGS, OBS and POS represent wild type, PPARγ2 KO, ob/ob and POKO mice respectively.

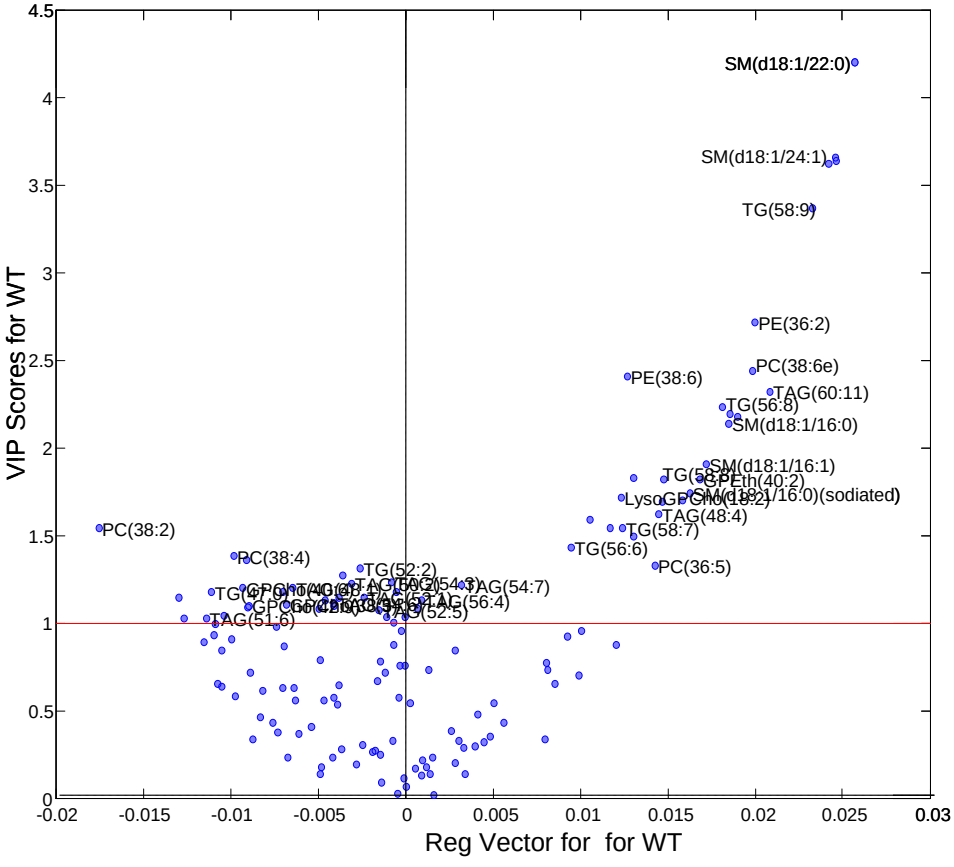


Figure 4.8. VIP analysis of serum lipidomic data for wild type group.

4.4.2 Lipidomic characterization of low and high HDL-C subjects

Low levels of high density lipoprotein cholesterol (HDL-C) is a recognised risk factor for heart disease. However, the mechanisms of how low HDL-C is contributed to cardiovascular diseases are still unknown. Here, lipidomics was utilised to study HDL derived from well characterized high and low HDL-C subjects**.** The study comprised 47 subjects: 24 low-HDL subjects and 23 high-HDL subjects who were participants of the Health 2000 Health Examination Survey. The subjects represented the extreme ends of HDL-C levels ($\leq 10^{th}$ and $\geq 90^{th}$ percentiles) and the HDL-C limits were as follows: for low-HDL-C men $\leq$ 1.03 mmol/l, low-HDL-C women $\leq$ 1.23 mmol/l, high-HDL-C men $\geq$ 1.79 mmol/l, and high-HDL-C women $\geq$ 2.24 mmol/l. Subjects with diabetes, alcohol abuse,

or malignancy were excluded. Alcohol abuse was defined as >160 grams of alcohol / week for women and >310 grams of alcohol / week for men.

Non-targeted lipidomic analysis was performed on low and high HDL-C subjects (Publication **III**) to uncover the differences in lipid composition and the resulting information was used to reconstitute HDL particles computationally.

Univariate analysis from both clinical measurements and lipidomic profiles was performed in order to find the parameters characterising low and high HDL-C subjects. Clinical and biochemical characteristics of low and high HDL-C subjects are summarised in (Table I, Publication **III**) and lipid species selected based on *p*-values ($p < 0.0001$) from student *t*-test between high and low HDL-C subjects are summarised in Table S1 of Publication **III**. Box plots for the most abundant lipids from lysoPC, SM, ChoE, ethanolamine plasmalogen (PEp) and TG classes are shown in Figure 4.10.

Supervised classification model was built for clustering and discrimination using partial least squares discriminant analysis (PLS/DA). The random subsets cross validation method and Q^2 scores were used to optimise the models. Based on the cross-validation, the model with two latent variables and $Q^2 = 0.51$ was selected. PLS/DA scores plot revealed clear separation between the two HDL-C groups (Figure 4.9A). Additionally, VIP analysis was performed to discover lipid variables responsible for the observed separation. Identified lipids with VIP value greater than two were further explored using fold changes and hierarchical clustering analysis. Heat map of fold changes with both samples and lipid variables ordered by hierarchical clustering is shown in Figure 4.9B. The clustering was based on Euclidean distance measure. We found that top VIP lipid variables from PLS/DA model formed two predominant clusters: mostly low HDL-C subjects and mostly high HDL-C subjects. Few mis-clustered samples may partly be attributed to individual variability. The fold changes in heat map were reflecting lipid profile changes relative to the average intensity of lipids within low HDL-C subjects. Bar plot in Figure 4.9B shows the mean fold change value of top VIP lipids within HDL-C subjects as compared to low HDL-C subjects.

Linear association of top VIP lipid variables with measured clinical variables were investigated using correlation analysis. Pearson correlations were computed between clinical and lipid variables. Both clinical and lipid variables were clustered using hierarchical clustering (Figure S3, Publication **III**). Interesting correlations were observed between lipid variables and HDL-C clinical parameters. Concentrations of TG molecular species were negatively

correlated with HDL-C concentration, while the SM and lysoPC lipid species were positively correlated with HDL-C parameter. However, PCs did not show any general trend in correlation with HDL-C parameter. We further investigated HDL-C concentration with some top lipid variables in both low and high HDL-C subjects (Figure S4 and S5, Publication **III**). Positive correlations between SM(d18:1/16:0) and HDL-C were similar in low HDL-C subjects ($r = 0.71$, $p = 0.0001$) and high HDL-C subjects ($r = 0.71$, $p = 0.0001$). Interestingly, positive correlation between lysoPC(18:0) and HDL-C in low HDL-C subjects ($r = 0.54$, $p = 0.006$) disappeared in high HDL-C subjects ($r = -0.06$, $p = 0.78$). We observed no correlation of TG(16:0/18:1/20:1) with HDL-C in low HDL-C subjects ($r = 0.06$, $p = 0.75$) and negative correlation in high HDL-C subjects ($r = -0.43$, $p = 0.04$). Additionally, linear association was also investigated for ChoEs with HDL-C. No significant correlations were found with concentrations for ChoE(18:1) and HDL-C variable either in low HDL-C subjects ($r = 0.33$, $p = 0.12$) or high HDL-C subjects ($r = 0.02$, $p = 0.92$).

The lipidomic level studies enabled to have a closer look at the molecular level details which are used in the simulation studies. While TLC or HPLC methods may serve as a valuable tool to analyse class-specific changes, lipidomics analysis not only guided to reconstitute HDL particle using simulation studies but also enabled us to investigate lipid molecular composition (Publication III). All previous simulations in the field have been based on either a single-component lipid particle composed of phosphatidylcholines lipids, or a two-component mixture of phosphatidylcholines and cholesteryl esters.

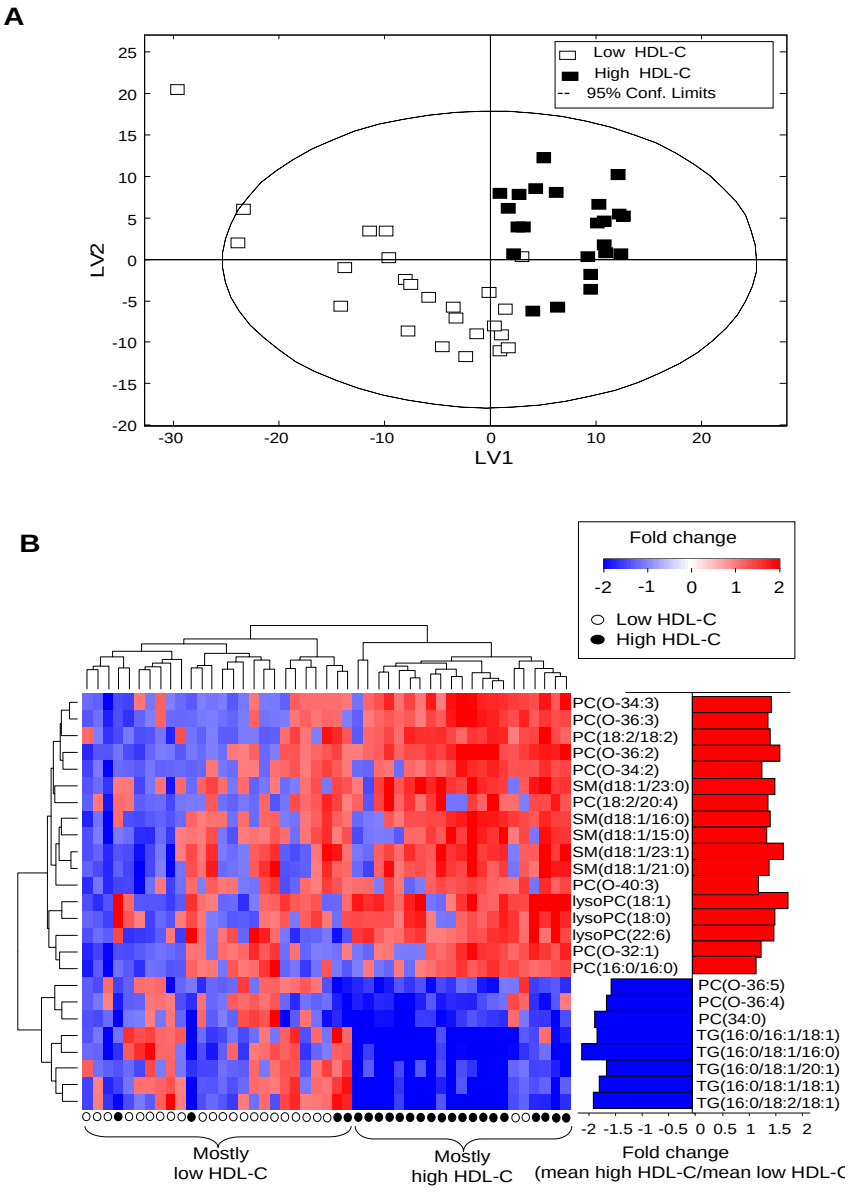


Figure 4.9 A. Partial least squares discriminant analysis (PLS/DA) of lipidomic profiles for low HDL-C and high HDL-C subjects. PLS/DA scores plot with two different HDL-C groups are indicated. B. Hierarchical clustering on most important VIP variables and samples in the heat map reflecting fold changes of lipids relative to mean intensity within low HDL-C group.

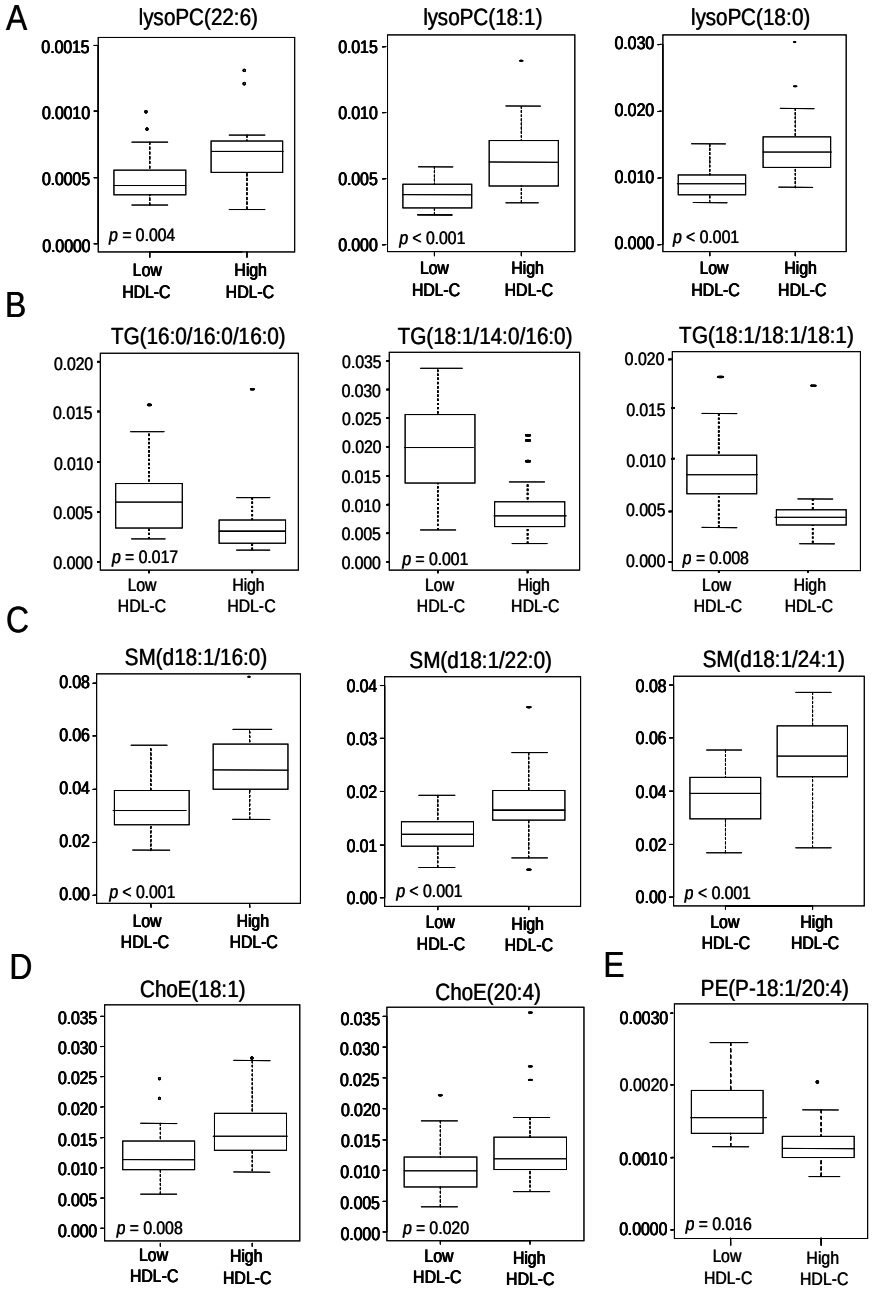


Figure 4.10. Box plots of the most abundant lipids within the TG, lysoPC, SM, ChoE and ethanolamine plasmalogen (PEp) classes. Concentrations are shown in mmol/l [lipid] / mg/dl [apoA-I].

5. SUMMARY AND CONCLUSIONS

Modern MS-based analytical technologies have generated wealth of information in lipidomics studies. The information presented many challenges for bioinformaticians due to its complexity. The goal of this thesis was to address such challenges in the context of non-targeted lipidomics studies based on UPLC/MS experimental methodologies.

A bioinformatics approach for the construction of lipid database for major classes of lipids is presented. Each lipid entry in the database was assigned with relevant information such as lipid names including short names, SMILES, scores, molecular weight, monoisotopic mass, isotope distribution (Publication **I**). The database was customised for UPLC/MS experiments by incorporating the information such as retention time range, adducts and main fragments to screen for potential lipids. This database information facilitated greatly building of experimental tandem MS libraries for different biological tissues. An example of such spectral libraries was built for different lipoprotein fractions (Publication **VI**).

Non-targeted metabolomic studies are often challenged by the presence of unknowns and hence present an additional challenge for the interpretation of experimental data. Frequently used supervised classification methods were employed for the functional prediction of class labels for unknown lipids to facilitate exploratory analysis as well as simplify the identification process (Publication **II**). As lipidomics goes beyond detecting the complete inventory of lipids, a new strategy called pathways *instantiation* is proposed to understand lipids in the context of pathways and thereby providing insights for the phenotype characterization (Publication **I**).

Lipid profiling was successfully applied to study mouse models in the context of POKO mice (Publication **IV** and **V**) and humans in the context of high and low HDL-C subjects (Publication **III**). Chemometric methods such as principal component analysis (PCA) and partial least squares and discriminant analysis

(PLS/DA) were employed for exploratory analysis as well as biomarker discovery in the context of different phenotypes such as characterization of high and low HDL-C subjects (Publication **III**) and POKO mice (Publication **IV**).

5.1 Future perspectives

While bioinformatics strategies presented herein facilitate lipidomic studies, the existing lipid informatics methods need to be extened to include more complex lipids and their mass spectra to databases to facilitate screening lipids such as steroids and glycolipids. The computational methodologies can be developed for the classification of unidentified metabolites for other analytical platforms. The novel tools are expected to integrate lipidome data with other omics level information in a context-dependent manner and thereby establishing lipid networks for underlying phenotypes. The complexity of lipidomes and their regulation at multiple levels makes their study a challenge for bioinformaticians and computational biologists.

REFERENCES

Anisimova, M., and Z. Yang. 2007. Multiple Hypothesis Testing to Detect Lineages under Positive Selection that Affects Only a Few Sites. *Mol. Biol. Evol.* 24(5): 1219–1228.

Apollonio, L. G., D. J. Pianca, I. R. Whittall, W. A. Maher, and J. M. Kyd. 2006. A demonstration of the use of ultra-performance liquid chromatography-mass spectrometry [UPLC/MS] in the determination of amphetamine-type substances and ketamine for forensic and toxicological analysis. *J. Chromatogr. B* 836(1–2): 111–115.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* 57(1): 289–300.

Bennet, R. D., and E. Heftmann. 1962. Thin-layer chromatography of sterols. *J. Chromatogr.* 9(9): 359–62.

Bijlsma, S., I. Bobeldijk, E. R. Verheij, R. Ramaker, S. Kochhar, I. A. Macdonald, B. van Ommen, and A. K. Smilde. 2005. Large-Scale Human Metabolomics Studies: A Strategy for Data (Pre-) Processing and Validation. *Anal.Chem.* 78(2): 567–574.

Brown, M. P. S., W. N. Grundy, D. Lin, N. Cristianini, C. W. Sugnet, T. S. Furey, M. Ares, and D. Haussler. 2000. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc.Natl. Acad. Sci. USA 97(1): 262–267.

Brugger, B., G. Erben, R. Sandhoff, F. T. Wieland, and W. D. Lehmann. 1997. Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem masspectrometry. *Proc. Natl. Acad.Sci. USA* 94(6): 2339–2344.

Caffrey, M., and J. Hogan. 1992. LIPIDAT: A database of lipid phase transition temperatures and enthalpy changes. DMPC data subset analysis. *Chem. Phys. Lipids* 61(1): 1–109.

Cai, Y.-D., X.-J. Liu, X.-b. Xu, and G.-P. Zhou. 2001. Support Vector Machines for predicting protein structural class. *BMC Bioinformatics* 2(1): 3.

Churchwell, M. I., N. C. Twaddle, L. R. Meeker, and D. R. Doerge. 2005. Improving LC-MS sensitivity through increases in chromatographic performance: Comparisons of UPLC-ES/MS/MS to HPLC-ES/MS/MS. *J. Chromatogr. B* 825(2): 134–143.

Cotter, D., A. Maer, C. Guda, B. Saunders, and S. Subramaniam. 2006. LMPD: LIPID MAPS proteome database. *Nucl. Acids Res.* 34(suppl_1): D507–510.

Cutler, R. G., J. Kelly, K. Storie, W. A. Pedersen, A. Tammara, K. Hatanpää, J. C. Troncoso, and M. P. Mattson. 2004. Involvement of oxidative stress-induced abnormalities in ceramide and cholesterol metabolism in brain aging and Alzheimer's disease. *Proc.Natl. Acad. Sci. USA* 101(7): 2070–2075.

Druilhet, R. E., M. L. Overturf, and W. M. Kirkendall. 1975. Structure of neutral glycerides and phosphoglycerides of human kidney. *Int. J. Biochem.* 6: 893–901.

Duda, R. O., P. E. Hart, and D. G. Stork. 2001. *Pattern Classification*, John Wiley & Sons, New York.

Ejsing, C. S., E. Duchoslav, J. Sampaio, K. Simons, R. Bonner, C. Thiele, K. Ekroos, and A. Shevchenko. 2006. Automated Identification and Quantification of Glycerophospholipid Molecular Species by Multiple Precursor Ion Scanning. *Anal. Chem.* 78(17): 6202–6214.

Ejsing C.S., J.L. Sampaio, V. Surendranath, E. Duchoslav, K. Ekroos, et al. 2009. Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *PNAS* 106(7): 2136–2141.

Ekroos, K., I. V. Chernushevich, K. Simons, and A. Shevchenko. 2002. Quantitative profiling of phospholipids by multiple precursor ion scanning on a hybrid quadrupole time-of-flight mass spectrometer. *Anal. Chem.* 74(5): 941–949.

Fahy, E., S. Subramaniam, H. A. Brown, C. K. Glass, A. H. Merrill, Jr., R. C. Murphy, C. R. H. Raetz, D. W. Russell, Y. Seyama, W. Shaw, T. Shimizu, F. Spener, G. van Meer, M. S. VanNieuwenhze, S. H. White, J. L. Witztum, and E. A. Dennis. 2005. A comprehensive classification system for lipids. *J. Lipid Res.* 46(5): 839–862.

Fahy, E., S. Subramaniam, R. C. Murphy, M. Nishijima, C. R. H. Raetz, T. Shimizu, F. Spener, G. van Meer, M. J. O. Wakelam, and E. A. Dennis. 2009. Update of the LIPID MAPS comprehensive classification system for lipids. J. Lipid Res. 50(Supplement): S9–14.

Fahy, E., M. Sud, D. Cotter, and S. Subramaniam. 2007. LIPID MAPS online tools for lipid research. *Nucl. Acids Res.* 35(suppl_2): W606–612.

Farcomeni, A. 2008. A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Stat. Methods Med. Res.* 17(4): 347–388.

Fernandis, A. Z., and M. R. Wenk. 2009. Lipid-based biomarkers for cancer. J. Chromatogr*. B* 877(26): 2830–2835.

Fisher Box, J. 1987. Guinness, Gosset, Fisher, and Small Samples. Stat. Sci. 2(1): 45–52.

Gopalacharyulu, P. V., E. Lindfors, C. Bounsaythip, T. Kivioja, L. Yetukuri, J. Hollmén, and M. Orešič. 2005. Data integration and visualization system for enabling conceptual biology. *Bioinformatics* 21(Supplement 1): i177–185.

Gray, S., and A. Vidal-Puig. 2007. Adipose Tissue Expandability in the Maintenance of Metabolic Homeostasis. *Nutr. Rev.* 65(s1): S7–S12.

Griffiths, W. J. 2003. Tandem mass spectrometry in the study of fatty acids, bile acids, and steroids. *Mass Spectrom. Rev.* 22(2): 81–152.

Gupta, S., M. R. Maurya, D. L. Stephens, E. A. Dennis, and S. Subramaniam. 2009. An Integrated Model of Eicosanoid Metabolism and Signaling Based on Lipidomics Flux Analysis. *Biophys. J.* 96(11): 4542–4551.

Hackl, H., M. Maurer, B. Mlecnik, J. Hartler, G. Stocker, D. Miranda-Saavedra, and Z. Trajanoski. 2004. GOLD.db: genomics of lipid-associated disorders database. *BMC Genomics* 5(1): 93–98.

Han, X., and R. W. Gross. 2001. Quantitative Analysis and Molecular Species Fingerprinting of Triacylglyceride Molecular Species Directly from Lipid Extracts of Biological Samples by Electrospray Ionization Tandem Mass Spectrometry. *Anal. Biochem.* 295(1): 88–100.

Han, X., and R. W. Gross. 2005a. Shotgun lipidomics: Electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. *Mass Spec. Rev.* 24(3): 367–412.

Han X, Gross RW. 2005b. Shotgun lipidomics: multidimensional MS analysis of cellular lipidomes. *Expert Review of Proteomics* 2: 253–64

Hand, D. J., H. Mannila, and P. Smyth. 2001. *Principles of Data Mining*. MIT Press.

Haynes, C. A., J. C. Allegood, H. Park, and M. C. Sullards. 2009. Sphingolipidomics: Methods for the comprehensive analysis of sphingolipids. *J.Chromatogr. B* 877(26): 2696–2708.

Hermansson, M., A. Uphoff, R. Kakela, and P. Somerharju. 2005. Automated Quantitative Analysis of Complex Lipidomes by Liquid Chromatography/Mass Spectrometry. *Anal. Chem.* 77(7): 2166–2175.

Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24: 417–441.

Houjou, T., K. Yamatani, M. Imagawa, T. Shimizu, and R. Taguchi. 2005. A shotgun tandem mass spectrometric analysis of phospholipids with normal-phase and/or reverse-phase liquid chromatography/electrospray ionization mass spectrometry. *Rapid Comm. Mass Spectrom.* 19(5): 654–666.

Joachims, T. 1998. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *Proc. ECML-98, 10th European Conference on Machine Learning*: 137–142.

Jolliffe, I. T. 1986. *Principal Component Analysis*. Springer-Verlag, New York.

Joyce, A. R., and B. O. Palsson. 2006. The model organism as a system: integrating 'omics' data sets. *Nat. Rev. Mol. Cell. Biol.* 7(3): 198–210.

Junker, B., C. Klukas, and F. Schreiber. 2006. VANTED: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics* 7(1): 109.

Kaluzny, M., L. Duncan, M. Merritt, and D. Epps. 1985. Rapid separation of lipid classes in high yield and purity using bonded phase columns. *J. Lipid Res.* 26(1): 135–140.

Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl. Acids Res.* 28(1): 27–30.

Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, and M. Hattori. 2004. The KEGG resource for deciphering the genome. *Nucl. Acids Res.* 32(Database issue): D277–280.

Katajamaa, M., J. Miettinen, and M. Orešič. 2006. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* 22(5): 634–636.

Katajamaa, M., and M. Orešič. 2007. Data processing for mass spectrometry-based metabolomics. *J. Chromatogr. A* 1158(1–2): 318–328.

Katajamaa, M., and M. Orešič. 2005. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics* 6(1): 179–190.

Keseler, I. M., J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I. T. Paulsen, M. Peralta-Gil, and P. D. Karp. 2005. EcoCyc: a comprehensive database resource for Escherichia coli. *Nucl. Acids Res.* 33(suppl_1): D334–337.

Khaselev, N., and R. C. Murphy. 2000. Structural characterization of oxidized phospholipid products derived from arachidonate-containing plasmenyl glycerophosphocholine. *J. Lipid Res.* 41(3): 564–572.

Koutnikova, H., T.-A. Cock, M. Watanabe, S. M. Houten, M.-F. Champy, A. Dierich, and J. Auwerx. 2003. Compensation by the muscle limits the metabolic consequences of lipodystrophy in PPAR gamma hypomorphic mice. *Proc. Natl. Acad. Sci. USA* 100(24): 14457–62.

Krieger, C. J., P. Zhang, L. A. Mueller, A. Wang, S. Paley, M. Arnaud, J. Pick, S. Y. Rhee, and P. D. Karp. 2004. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucl. Acids Res.* 32(suppl_1): D438–442.

Leandro, C. C., P. Hancock, R. J. Fussell, and B. J. Keely. 2006. Comparison of ultra-performance liquid chromatography and high-performance liquid chromatography for the determination of priority pesticides in baby foods by tandem quadrupole mass spectrometry. *J. Chromatogr. A* 1103(1): 94–101.

Leavell, M. D., and J. A. Leary. 2006. Fatty Acid Analysis Tool (FAAT): An FT-ICR MS Lipid Analysis Algorithm. *Anal. Chem.* 78(15): 5497–5503.

Lesnefsky, E. J., M. S. K. Stoll, P. E. Minkler, and C. L. Hoppel. 2000. Separation and Quantitation of Phospholipids and Lysophospholipids by High-Performance Liquid Chromatography. *Anal. Biochem.* 285(2): 246–254.

Lu, Y., S. Hong, and C. Serhan. 2006. Lipid Mediator Informatics-Lipidomics: Novel Pathways in Mapping Resolution. *AAPS Journal* 8(2): E284–E297.

Lusis, A. J. 2000. Atherosclerosis. *Nature* 407(6801): 233–241.

Mann, H. B., and D. R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* 18: 50–60.

Matthew Barker, W. R. 2003. Partial least squares for discrimination. *J. Chemometr.* 17(3): 166–173.

McHowat, J., J. H. Jones, and M. H. Creer. 1997. Gradient elution reversed-phase chromatographic isolation of individual glycerophospholipid molecular species. *J.Chromatogr. B* 702(1–2): 21–32.

Menendez, J. A., and R. Lupu. 2007. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat. Rev. Cancer* 7(10): 763–777.

Merrill, J., Alfred H., M. C. Sullards, J. C. Allegood, S. Kelly, and E. Wang. 2005. Sphingolipidomics: High-throughput, structure-specific, and quantitative analysis of sphingolipids by liquid chromatography tandem mass spectrometry. *Methods* 36(2): 207–224.

Michalec, C., M. Sulc, and J. Mestan. 1962. Analysis of Cholesteryl Esters and Triglycerides by Thin-Layer Chromatography. *Nature* 193: 63–64.

Miller, R. G. 1981. *Simultaneous statistical inference*. Springer Verlag, New York.

Mitchell, T. 1997. *Machine Learning*. McGraw Hill, New York.

Myoung Hee Lee, G. H. L., Jong Shin Yoo,. 2003. Analysis of ceramides in cosmetics by reversed-phase liquid chromatography/electrospray ionization mass spectrometry with collision-induced dissociation. *Rapid Comm. Mass Spectrom.* 17(1): 64–75.

Navas-Iglesias, N., A. Carrasco-Pancorbo, and L. Cuadros-Rodríguez. 2009. From lipids analysis towards lipidomics, a new challenge for the analytical chemistry of the 21st century. Part II: Analytical lipidomics. Trac-trend Anal. Chem. 28(14): 393–403.

Ogiso, H., T. Suzuki, and R. Taguchi. 2008. Development of a reverse-phase liquid chromatography electrospray ionization mass spectrometry method for lipidomics, improving detection of phosphatidic acid and phosphatidylserine. *Anal. Biochem.* 375(1): 124–131.

Orešič, M., V. A. Hänninen, and A. Vidal-Puig. 2008. Lipidomics: a new window to biomedical frontiers. *Trends Biotechnol.* 26(12): 647–652.

Panganamala, R. V., L. A. Horrocks, J. C. Geer, and D. G. Cornwell. 1971. Positions of double bonds in the monounsaturated alk-1-enyl groups from the plasmalogens of human heart and brain. *Chem. Phys. Lipids* 6(2): 97–102.

Pearson, K. 1896. Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philos. Trans. Royal Soc.* 187: 253–318.

Pietiläinen, K. H., M. Sysi-Aho, A. Rissanen, T. Seppänen-Laakso, H. Yki-Järvinen, J. Kaprio, and M. Orešič. 2007. Acquired Obesity Is Associated with Changes in the Serum Lipidomic Profile Independent of Genetic Effects- A Monozygotic Twin Study. *PLoS ONE* 2(2): e218.

Pulfer, M., and R. C. Murphy. 2003. Electrospray mass spectrometry of phospholipids. *Mass Spectrom. Rev.* 22(5): 332–364.

Rosen, E. D., P. Sarraf, A. E. Troy, G. Bradwin, K. Moore, D. S. Milstone, B. M. Spiegelman, and R. M. Mortensen. 1999. PPAR[gamma] Is Required for the Differentiation of Adipose Tissue In Vivo and In Vitro. *Mol. Cell* 4(4): 611–617.

Schwudke, D., J. Oegema, L. Burton, E. Entchev, J. T. Hannich, C. S. Ejsing, T. Kurzchalia, and A. Shevchenko. 2005. Lipid Profiling by Multiple Precursor and Neutral Loss Scanning Driven by the Data-Dependent Acquisition. Anal.Chem. 78(2): 585–595.

Shi, Y., and P. Burn. 2004. Lipid metabolic enzymes: emerging drug targets for the treatment of obesity. *Nat. Rev. Drug Discov.* 3(8): 695–710.

Snedecor, G. W., and W. G. Cochran. 1989. Statistical Methods, Iowa State University Press.

Snyder, F. 1999. The ether lipid trail: a historical perspective. *Biochimica et Biophysica Acta (BBA) – Molecular and Cell Biology of Lipids* 1436(3): 265–278.

Song, H., F.-F. Hsu, J. Ladenson, and J. Turk. 2007. Algorithm for Processing Raw Mass Spectrometric Data to Identify and Quantitate Complex Lipid Molecular Species in Mixtures by Data-Dependent Scanning and Fragment Ion Database Searching. *J. Am. Soc. Mass Spectrom.* 18(10): 1848–1858.

Spearman, C. 1904. The proof and measurement of association between two things. Amer. J*. Psychol.* 15: 72–101.

Spiegelman, B. M. 1998. PPAR-gamma: adipogenic regulator and thiazolidinedione receptor. *Diabetes* 47(4): 507–514.

Ståhlman, M., C. S. Ejsing, K. Tarasov, J. Perman, J. Borén, and K. Ekroos. 2009. High-throughput shotgun lipidomics by quadrupole time-of-flight mass spectrometry. *J. Chromatogr. B* 877(26): 2664–2672.

Sud, M., E. Fahy, D. Cotter, A. Brown, E. A. Dennis, C. K. Glass, A. H. Merrill, Jr, R. C. Murphy, C. R. H. Raetz, D. W. Russell, and S. Subramaniam. 2007. LMSD: LIPID MAPS structure database. *Nucl. Acids Res.* 35(suppl_1): D527–532.

Taskinen, M. R., T. Kuusi, E. Helve, E. A. Nikkila, and H. Yki-Järvinen. 1988. Insulin therapy induces antiatherogenic changes of serum lipoproteins in noninsulin-dependent diabetes. *Arteriosclerosis* 8(2): 168–177.

Thomas M.C., T.W. Mitchell, D.G. Harman, J.M. Deeley, J.R. Nealon, and S.J. Blanksby 2007. Ozone-Induced Dissociation: Elucidation of Double Bond Position within Mass-Selected Lipid Ions. *Analytical Chemistry* 80: 303–11

van Meer, G. 2005. Cellular lipidomics. *EMBO J.* 24(18): 3159–3165.

Vance, D. E., and J. E. Vance. 2008. *Biochemistry of lipids, lipoproteins and membranes.* Elsevier, Hungary.

Vapnik, V. 1995. *The nature of statistical learning theory*. Springer-Verlag, New York.

Wang, C., H. Kong, Y. Guan, J. Yang, J. Gu, S. Yang, and G. Xu. 2005. Plasma Phospholipid Metabolic Profiling and Biomarkers of Type 2 Diabetes Mellitus Based on High-Performance Liq uid Chromatography/Electrospray Mass Spectrometry and Multivariate Statistical Analysis. *Anal. Chem.* 77(13): 4108–4116.

Watson, A. D. 2006. Thematic review series: Systems Biology Approaches to Metabolic and Cardiovascular Disorders. Lipidomics: a global approach to lipid analysis in biological systems. *J. Lipid Res.* 47(10): 2101–2111.

Weininger, D. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inform. Comput. Sci. 28(1): 31–36.

Wenk, M. R. 2005. The emerging field of lipidomics. Nat. Rev. Drug Discov. 4(7): 594–610.

Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics Bull.* 1: 80–83.

Wilson, I. D., J. K. Nicholson, J. Castro-Perez, J. H. Granger, K. A. Johnson, B. W. Smith, and R. S. Plumb. 2005. High Resolution Ultra Performance Liquid Chromatography Coupled to oa-TOF Mass Spectrometry as a Tool for Differential Metabolic Pathway Profiling in Functional Genomic Studies. *J.Proteome Res.* 4(2): 591–598.

Wold, S., K. Esbensen, and P. Geladi. 1987. Principal Component analysis. *Chemometr. Intell. Lab. Syst.* 2: 37–52.

Yetukuri, L., K. Ekroos, A. Vidal-Puig, and M. Orešič. 2008. Informatics and computational strategies for the study of lipids. *Mol. BioSyst.* 4(2): 121–127.

Zemski Berry, K. A., and R. C. Murphy. 2004. Electrospray ionization tandem mass spectrometry of glycerophosphoethanolamine plasmalogen phospholipids. *J. Am. Soc. Mass Spectrom.* 15(10): 1499–1508.

Publications II and VI are not included in the PDF version. Please order the printed version to get the complete publication (http://www.vtt.fi/publications/index.jsp).

PUBLICATION I

Bioinformatics strategies for lipidomics analysis: characterization of obesity related hepatic steatosis

In: BMC Systems Biology 2007.
1:12.

BMC Systems Biology

BioMed Central

Research article

Bioinformatics strategies for lipidomics analysis: characterization of obesity related hepatic steatosis

Laxman Yetukuri[1], Mikko Katajamaa[2], Gema Medina-Gomez[3], Tuulikki Seppänen-Laakso[1], Antonio Vidal-Puig[3] and Matej Orešič*[1]

Address: [1]VTT Technical Research Centre of Finland, Tietotie 2, FIN-02044, Espoo, Finland, [2]Turku Centre for Biotechnology, Tykistökatu 6, FIN-20521, Turku, Finland and [3]University of Cambridge Department of Clinical Biochemistry, Addenbrooke's Hospital, Hills Road, CB2 2QR, Cambridge, UK

Email: Laxman Yetukuri - ext-laxman.yetukuri@vtt.fi; Mikko Katajamaa - mikko.katajamaa@btk.utu.fi; Gema Medina-Gomez - mgm28@cam.ac.uk; Tuulikki Seppänen-Laakso - tuulikki.seppanen-laakso@vtt.fi; Antonio Vidal-Puig - ajv22@medschl.cam.ac.uk; Matej Orešič* - matej.oresic@vtt.fi

* Corresponding author

Published: 15 February 2007

BMC Systems Biology 2007, **1**:12 doi:10.1186/1752-0509-1-12

Received: 7 November 2006
Accepted: 15 February 2007

This article is available from: http://www.biomedcentral.com/1752-0509/1/12

Abstract

Background: Lipids are an important and highly diverse class of molecules having structural, energy storage and signaling roles. Modern analytical technologies afford screening of many lipid molecular species in parallel. One of the biggest challenges of lipidomics is elucidation of important pathobiological phenomena from the integration of the large amounts of new data becoming available.

Results: We present computational and informatics approaches to study lipid molecular profiles in the context of known metabolic pathways and established pathophysiological responses, utilizing information obtained from modern analytical technologies. In order to facilitate identification of lipids, we compute the scaffold of theoretically possible lipids based on known lipid building blocks such as polar head groups and fatty acids. Each compound entry is linked to the available information on lipid pathways and contains the information that can be utilized for its automated identification from high-throughput UPLC/MS-based lipidomics experiments. The utility of our approach is demonstrated by its application to the lipidomic characterization of the fatty liver of the genetically obese insulin resistant ob/ob mouse model. We investigate the changes of correlation structure of the lipidome using multivariate analysis, as well as reconstruct the pathways for specific molecular species of interest using available lipidomic and gene expression data.

Conclusion: The methodology presented herein facilitates identification and interpretation of high-throughput lipidomics data. In the context of the ob/ob mouse liver profiling, we have identified the parallel associations between the elevated triacylglycerol levels and the ceramides, as well as the putative activated ceramide-synthesis pathways.

Background

Lipids are a diverse class of biological molecules that play a central role as structural components of biological membranes, energy reserves, and signaling molecules [1]. They are broadly defined as hydrophobic or amphipathic small molecules that may originate entirely or in part by carbanion based condensation of thioesters, and/or by carbocation based condensation of isoprene units [2]. Lipids also contribute to common pathophysiological states such as fatty liver and lipotoxic induced insulin resistance, Alzheimer's disease, atherosclerosis, and toxic manifestations of infectious diseases [3,4]. Therefore identification and characterization of these metabolic networks offers a unique opportunity to devise therapeutic strategies to prevent or reverse these pathological states.

While lipid-, and metabolome research in general, over past decades was overshadowed by the progress of genomics, recent revived and burgeoning interest in lipids that triggered several new endeavors in lipid research illustrates their critical biological importance. Lipidomics as a field aims at characterization of lipid molecular species and their biological roles with respect to the expression of proteins involved in lipid metabolism and function including gene regulation [5,6].

Several useful public resources exist representing various aspects of information on lipids, such as LIPID MAPS [7,8], Lipid Bank [9], CyberLipids [10], and LIPIDAT [11]. The LIPID MAPS consortium introduced a nomenclature that enables to represent a lipid compound by a unique 12-digit identifier [2]. LIPID MAPS also includes tandem mass spectrometry (MS/MS) fragment information for several lipid molecular species.

With the enhanced capabilities of modern MS instruments and interfaces, there has been an increase in development of global lipid analytical methods, either using liquid chromatography mass spectrometry (LC/MS) based methods focused on sensitive analyses of total lipid extracts or on specific classes of metabolites [12-15], or direct MS^n scanning driven by data-dependent acquisition [16-19] without chromatographic separation. Due to the structural characteristics of lipids their identification from fragment mass spectra is generally easier than for other molecular components and today's typical global lipid profiling analyses allows identifying of several hundred lipid molecular species in parallel. Informatics strategies have already been developed which utilize mass spectrometry based approaches in combination with database searches to rapidly identify specific classes of lipids, such as phospholipids [16] or PUFA-derived lipid mediators [20]. While much further progress is still needed in the area of lipid analytics, one of the biggest challenges is elucidation of biological phenomena behind the large amounts of lipidomics data currently available.

Advances in analytical methods, along with improved data processing software solutions [21-25], demand development of comprehensive lipid libraries to allow system level identification, discovery, and subsequent study of lipids. Integrative studies combining multi-tissue lipidomic profiles with other levels of biological information such as gene expression and proteomics have been made possible due to such capabilities [26,27]. Currently available databanks such as LIPID MAPS offer a necessary starting point for explorations of the lipidome and a reference for validation of results. However, in context of high-throughput lipidomic profiling and systems biology studies, the currently available online resources face threefold challenge:

1. Due to high volumes of information available from high-throughput lipidomics experiments, the database system has to be efficiently linked to the analytical platform generating the lipid profile data, as well as to chemo- and bioinformatics system for compound identification and linking the information to other levels of biological organization to enable systems approaches.

2. Due to diversity of lipids across different organisms, tissues, and cell types, it is unlikely any one database can cover all possible lipids. A mechanism is therefore necessary that facilitates identification as well as discovery of new lipid species in biological systems from available data.

3. Currently available pathway-level representation of lipids in databases such as KEGG [28] is limited to pathway representation of generic lipid classes, i.e. including mainly the head group information, and not including the fatty acid side chain information. Therefore, these lipid databases lack the level of detail that is becoming available by modern LC/MS based approaches.

Additionally, due to common structural features of different lipid classes, often regulated by the same enzymes in class-specific manner, there is a large degree of co-regulation to be expected in cellular, tissue, or biofluid lipid profiles. In order to elucidate the changes of the organism lipidome as a result of interventions, the data analysis and interpretation therefore needs to balance the analysis of global lipid pattern changes with the analysis of molecular species specific pathways.

In this paper we report a bioinformatics strategy for lipidomics analysis. We utilize the recently developed nomenclature of lipids [2] to generate a diverse scaffold of lipid compounds represented by the Simplified Molecular

Input Line Entry System (SMILES) representation [29,30]. Each compound entry is linked to the available information on lipid pathways and contains the information that can be utilized for automated identification from high-throughput LC/MS-based lipidomics experiments. We investigate the changes of correlation structure of the lipidome using multivariate analysis, as well as reconstruct the pathways for specific molecular instances of interest using available lipidomic and gene expression data.

We validate our approach by investigating the lipid profiles associated with hepatic steatosis observed in ob/ob mice. Our results indicate that obesity associated hepatic steatosis involves increased liver deposition of short chain triacylglycerol species associated with proportional increase of reactive ceramide lipid species. Of interest, the contribution of triacylglycerol molecular species is heterogenesous as indicated by the presence of a subset of long triacylglycerol species that does not contribute to the development of steatosis. We also provide evidence of specific dysregulation of ceramide synthesis pathways in steatosis and the influence of gender on the liver lipid composition.

Results and discussion

Lipid informatics

In this paper we primarily focus on studies of glycerophospholipids, sphingolipids, glycerolipids, and sterol esters. The main structural variants among these classes are variation within one or more fatty acid chains and the head group (see an example in Figure 1). In order to facilitate automated identification of lipids from lipidomics experiments, we used the structural rules of lipid molecular species to computationally generate a diverse set of lipids from "seed" fatty acids most likely to occur in living systems (Additional file 1 lists the seed fatty acids utilized in this paper). Our current choice of seeds reflects bias toward the mammalian cells, but the approach is general enough to afford suitable modifications depending on the area of interest.

The fatty acid seeds are expressed in terms of Simplified Molecular Input Line Entry System (SMILES), which is a human readable linear indexing system of atoms and bonds, dictated by specific syntax rules [29]. The modular nature of the lipid structure makes the SMILES representation very suitable for the task due to ease of algorithmic manipulation of lipid (sub)structures and their modifications. While in general multiple SMILES representations can exist for any given compound, canonical versions that enable unique SMILES representation are available. We utilize the Daylight canonical SMILES representation (Daylight, Chemical information system, Inc.). We generate a generic SMILES template for different classes of lipids and apply parsers for varying fatty acid chain lengths in order to create all possible compounds of that class in the given window of chosen fatty acid chain length. Systematic names complying with nomenclature of LIPID MAPS were generated algorithmically (Additional file 2 lists the lipid classes generated and their sizes in the database). Daylight SMILES Toolkit was tailored to get molecular weights and exact masses of compounds using elemental masses taken from literature [31].

Our approach is illustrated below using a systematic construction of glycerophospholipids classes as an example:

1. Construct generic SMILES template for glycerophospholipid class. SMILES template showing fatty acid seed variables at the sn-1 and sn-2 positions and head group at sn-3 position is:

"C(SMILES for fatty acid seed variable(R1))C(SMILES for fatty acid seed variable (R2))COP(=O)([O-])O-X)", where X represents SMILES for relevant part of head groups as shown in Figure 1.

2. Use corresponding systematic names against fatty acid seed SMILES to generate names using common name template:

"1-name of seed variable R1-2-name of seed variable R2-sn-glycero-3- name corresponding to X".

3. Convert SMILES into canonical SMILES.

4. From SMILES, obtain molecular formula and calculate molecular weight.

5. Obtain isotopic distribution of that compound and tailor it to the resolution of mass spectrometer.

The differences in length and degree of unsaturation in fatty acyl/alkyl chains lead to large diversity within each lipid class. When matching such database with the experimental lipidomics results, the searches thus inevitably result in large number of hits, both due to multiple close matches in mass as well as due to limitations of the analytical approach. In order to facilitate sifting through the multiple hits, we set up a heuristic scoring scheme based on seed fatty acid composition as described in Methods.

Lipidomics data processing and identification

Our lipid profiling platform is based on non-targeted analysis of total lipid extracts using Ultra Performance Liquid Chromatography (UPLC) coupled to quadrupole time of flight mass spectrometry. The platform characteristics are described in detailed elsewhere [32]. In order to better understand current limitations of the analytical strategy, as well as because our computational approaches

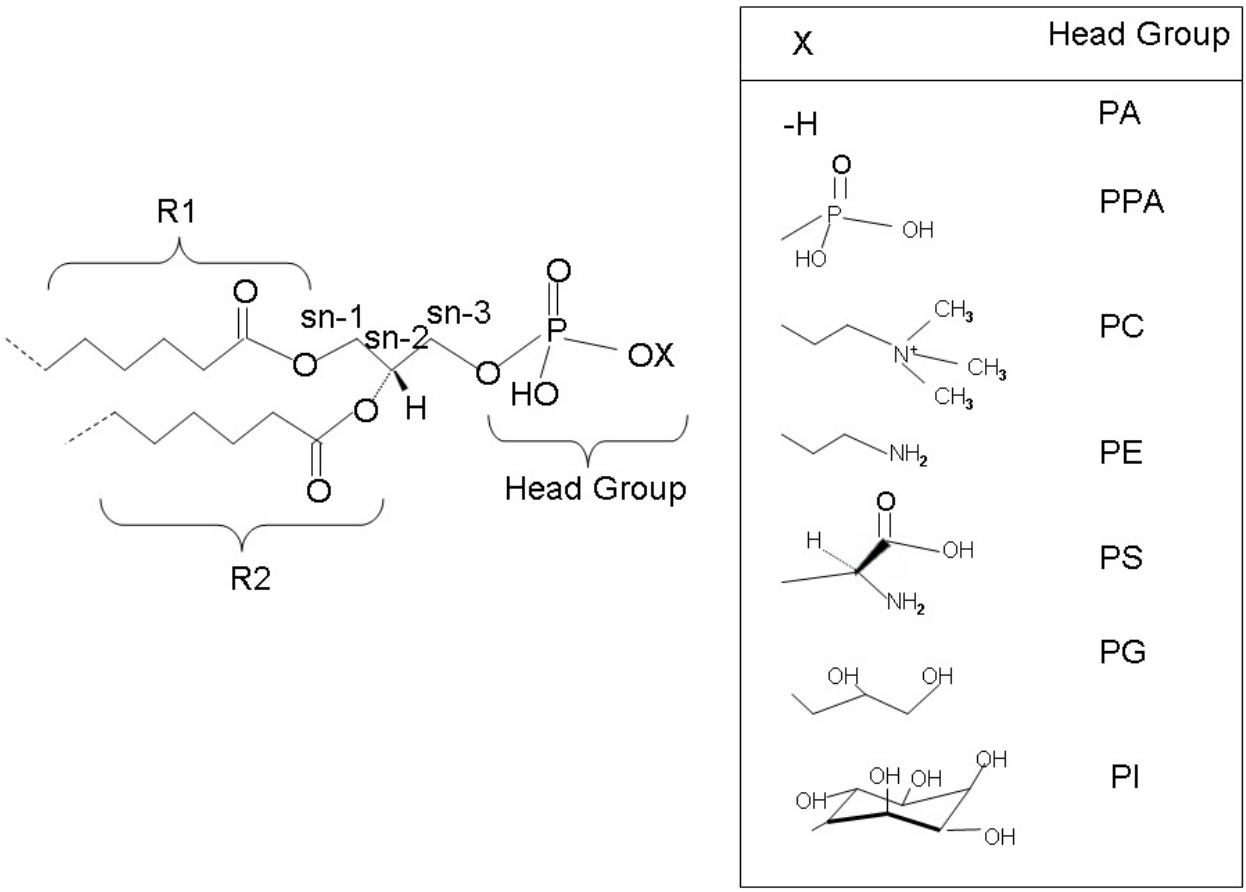


Figure 1
Structures of major glycerophospholipids. R1, R2 and X are SMILES seed variables at sn-1, sn-2 and sn-3 positions respectively. Head groups legend: PA = Phosphate, PPA = Pyrophosphate, PE = Phosphoethanolamine, PC = Phosphocholine, PS = Phosphoserine, PG = Phosphoglycerol, and PI = Phosphoinositol.

are adaptable to other platforms, including those using multiple precursor and neutral loss scanning [16,18], the analysis and data processing are described here only briefly.

An overview of the lipidomic data flow is shown in Figure 2. We convert raw mass spectrometer files to netCDF format to enable data processing with MZmine toolbox [21,22]. Peak detection and alignment parameters in MZmine are set based on preliminary investigation of platform specific characteristics such as peak shapes and retention time variation. Following the processing, each peak is characterized by mass-to-charge ratio (m/z) and retention time (RT) values.

In order to facilitate automated identification of lipids from peak lists, we compute the scaffold of theoretically possible lipids. LipidDB is a database of lipids constructed using SMILES, as described in the previous section. The internal library contains the platform-specific information about the internal standards and the lipid species identified using UPLC/MS/MS. To ease the identification of lipids based on the mass spectrometric data, we calculate isotopic distribution for every molecular species in both databases. The isotopic distribution is based on observed natural abundance of each element in the chemical formula [31]. Isotopic masses and abundances of given chemical composition are predicted using Isotope Pattern Calculator version 1.4 [33]. While the exact iso-

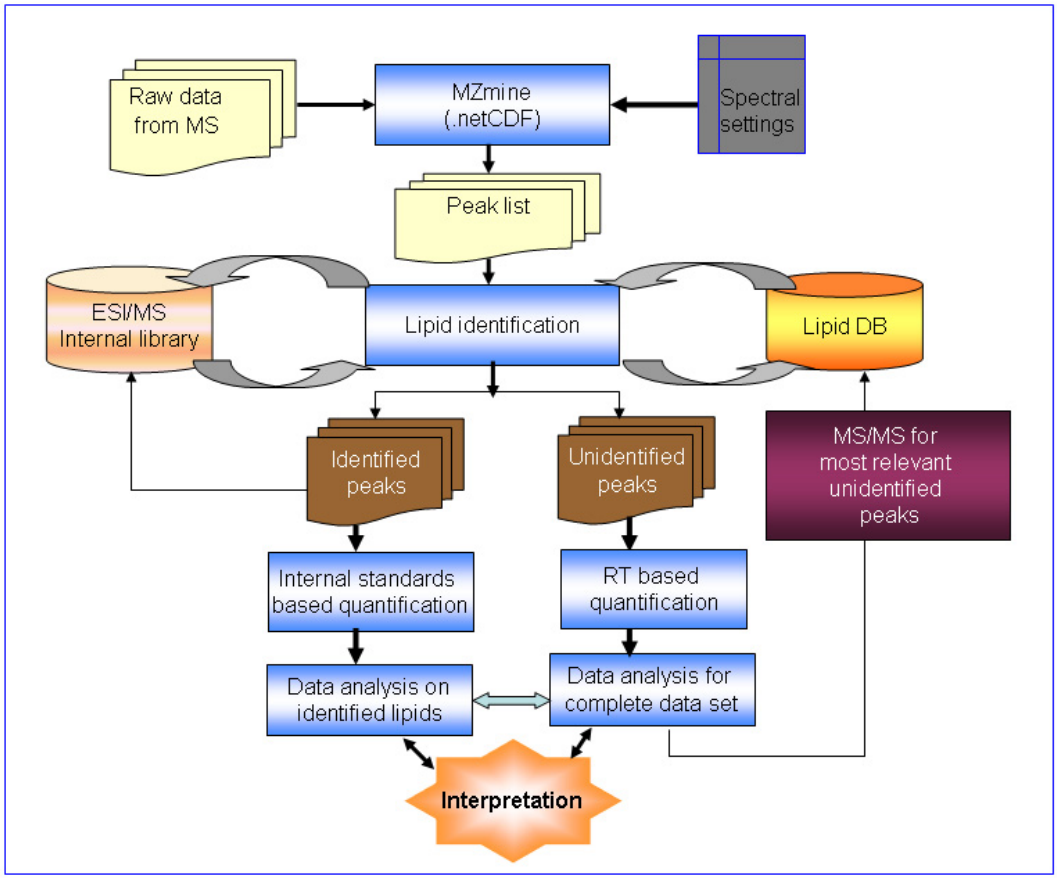


Figure 2
Lipidomic platform data flow. Summary of our lipidomic platform data flow from raw peak data to interpretation of spectra involving MZmine based data processing, lipid identification, quantification and multivariate data analysis.

tope patterns are kept in the database, the patterns are corrected for resolution of the mass spectrometer when matching with spectral data.

The internal library of lipids is searched first to ensure identification of internal standards and previously identified lipids. Retention times of these lipids are used as a constraint in lipid identification. The retention time information in part resolves the problem of identification of fatty acid moieties. The molecular species of the same class and carbon composition, but of different fatty acid composition, tend to elute at different times. The fatty acid composition can thus be determined in separate sample runs using tandem mass spectrometry (UPLC/MS/MS) in negative (phospho- or sphingolipids) or positive (acylglycerols) ion mode. In order not to compromise peak shapes in chromatographic direction, all reference UPLC/MS/MS spectra are generated in separate runs, which are set up so that ions selected for MS/MS analysis are well separated in elution time. We found the variation in retention times for the method described to be under 1.25%, as tested for multiple tissue or cell types over an extended period of time (over 18 months) for multiple UPLC C18 columns [32], therefore confirming retention time is a reliable parameter for the purpose of identification.

In the database, the redundancy due to varying fatty acid composition for the same molecular weight can be represented using the common notation showing total number of carbons and double bonds. For example, a diacylglycerophosphocholine species GPCho(16:0/

20:4(5Z,8Z,11Z,14Z)) (named as 1-hexadecanoyl-2-(5Z,8Z,11Z,14Z-eicosatetraenoyl)-sn-glycero-3-phosphocholine using LIPID MAPS nomenclature) could be represented also as GPCho(36:4). However, GPCho(36:4) could also represent other molecular species, for example GPCho(20:4(5Z,8Z,11Z,14Z)/16:0) or GPCho(18:2(9Z,12Z)/18:2(9Z,12Z)).

Peaks not identified by internal library are searched in LipidDB. Lipid identifications with LipidDB involve matching m/z, comparing RT range (based on knowledge on lipids from internal library), checking heuristic score and/or MS/MS. Matching of m/z value is a pre-requisite for identification. In some cases, isobaric species are distinguished based on retention time ranges and MS/MS. Protonated phosphocholine species are identified at even m/z and sphingomyelin species are identified at odd m/z values. We also check if identifications originate from the isotopic masses at the same retention time. Ultimately, identification of isobaric species, if not separated chromatographically, also depends on the mass resolution and type of the mass spectrometer. Specifically, we have observed co-fragmentation using UPLC/MS/MS in phosphatidylcholines and ethanolamine plasmalogens in a few instances. In such cases, instruments with MS^n capability and high resolution detectors (*i.e.*, Orbitrap or FTMS) may be necessary for exact identification.

Reconstruction of lipid molecular pathways

Following lipidomics data processing and identification, data analysis usually includes exploration of data as well as of their putative biological meaning. In addition to the level changes of specific metabolites, which can be analyzed using univariate statistical approaches, co-regulation of metabolites is also of interest. The interdependence of metabolites is driven by the underlying biophysical mechanisms such as chemical equilibrium, mass conservation, or asymmetric control distribution [34]. Since the lipids of the same class may be in part regulated by the same enzymes, high degree of within-class co-regulation is to be expected. Correlation network analysis has proved to be a valuable tool for exploring and visualizing co-regulations in metabolomics data [26,35,36]. A matrix of correlation coefficients, an indirect measure of distance between metabolites [37], is computed using pair-wise correlation between the corresponding concentrations of lipids in a given sample. The matrix is visualized in the form of metabolite correlation network based on a certain threshold criteria over correlation coefficient values.

In order to gain insight into the molecular mechanisms underlying the observed co-regulation (or similarly for de-regulation in specific context), the clustered lipids need to be mapped into the known pathways. Kyoto Encyclopedia of Genes and Genomes (KEGG) [22] has been the main source of information on metabolic pathways. However, KEGG lipid pathway representation is generally limited to generic lipid classes, *i.e.*, including mainly the head-group information, and not including the fatty acid side-chain information. As the level of information from MS studies is specific instance of subclass (*e.g.*, 1-octadecanoyl-2-dodecanoyl-sn-glycero-3-phosphocholine) and not the common sub class itself (*e.g.*, 1-acyl-2-acyl-sn-glycero-3-phosphocholine), a mechanism is necessary to convert generic enzymatic and pathway information from KEGG database to a specific instance under study. As we have implemented LIPID MAPS nomenclature, conversion of KEGG generic names into LIPID MAPS common subclass names and in turn to specific instance names allows mapping of identified lipids into pathways directly from MS-based studies with other levels of information.

We solve the limitation of generic lipid pathways by instantiating KEGG (or related) pathways for specific lipid molecular species of interest (Figure 3). In practice, our strategy to represent KEGG pathways involves the following steps:

1. Convert generic names of lipids in the KEGG reference lipid pathway into systematic common subclass names which enable to convert into systematic name for particular lipid as per LIPID MAPS consortium.

2. Construct XML schema to represent lipid pathway with systematic names of lipids and known EC numbers.

3. Generate XML document for a queried lipid.

4. Use megNet pathway visualization tool [38] to display the correlation network of lipids linked to pathways and ontologies.

Such approach affords visualization of pathways of interest in the context of observed biological data, including data from other levels such as microarray experiments. Presently we have not added additional level of quantitative analysis based on instantiated pathway information, but this is one of future considerations. One should bear in mind the complexity of such challenge as lipids are regulated systemically and their levels reflect complex systemic balance, therefore their pathways generally involve multiple tissues and complex dynamics [39].

Lipid profiling of liver tissue in an obese mouse model

We illustrate the combined informatics and analytical approach on the liver of ob/ob mice. The ob/ob is an obese, insulin resistant mouse model resulting from the spontaneous mutation of the *ob* gene encoding the leptin protein [40]. The ob/ob mouse is commonly studied as a

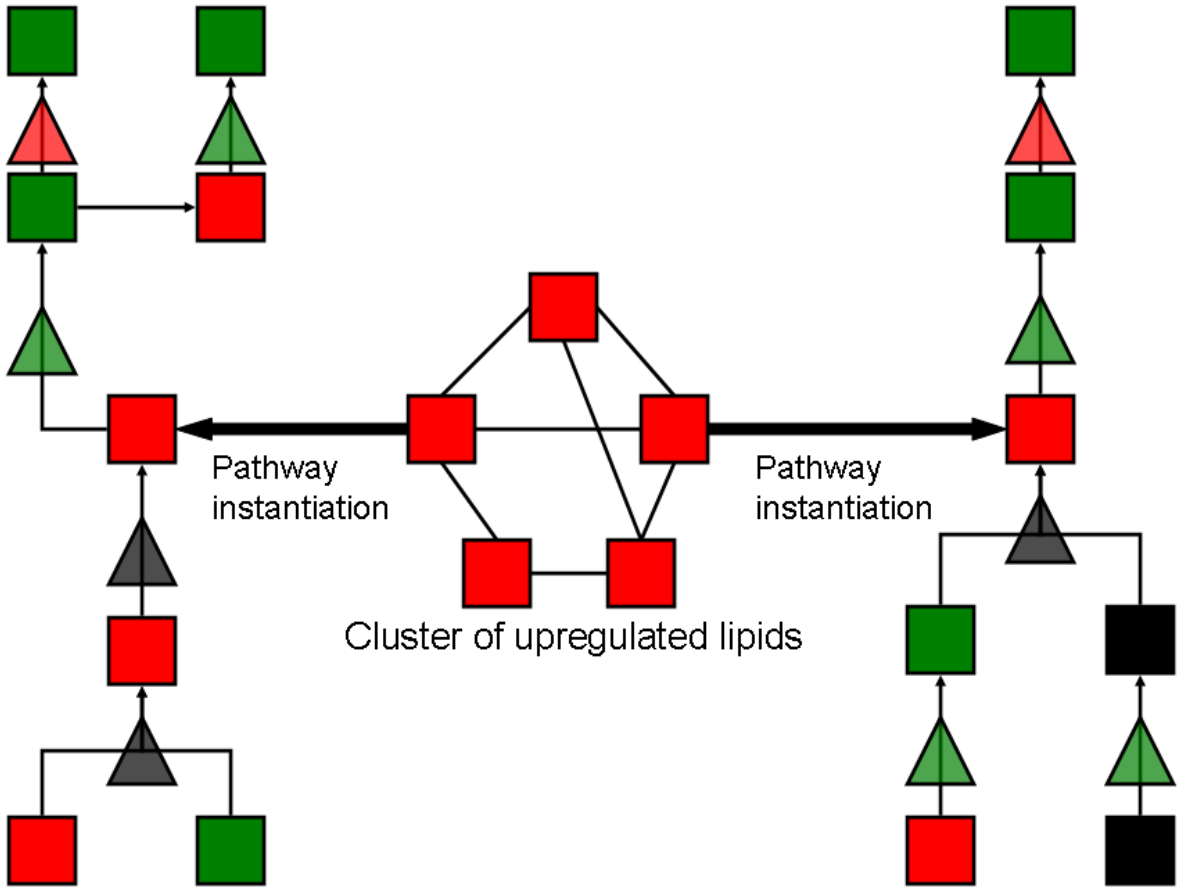


Figure 3
Lipid pathway instantiation. An illustrative example of instantiation of two co-regulated molecular species in the context of known lipid pathways. Upregulated species and enzymes (triangles) are in red, downregulated in green, unchanged in black.

model for early onset of severe obesity, insulin resistance and fatty liver. Figure 4 shows typical liver tissue images of ob/ob and wild type (WT) mouse, respectively. Lipid droplet accumulation is clearly seen in the obese model. Obesity is associated with deposition of triacylglycerols (TGs) in the liver tissue (hepatosteatosis). Fatty liver develops as a result of increased free fatty acid (FFA) availability in the context of obesity and insulin resistance associated to increased hepatic glucose production [1]. Elevated hepatic FFA levels, which further lead to increased esterification into TGs, may result from the combined effect of increased influx of plasma FFAs, increased *de novo* FFAs, and decreased β-oxidation [41].

The following genotypes were used for analysis: Wild Type (WT) and ob/ob. The study included 12 ob/ob (6 male, 6 female) and 10 WT (7 male, 3 female) mice of 16-week age. Figure 5 lists the results of ULPC/MS lipidomic profiling for selected molecular species, out of total 192 identified molecular species. Notable changes are upregulation in the ob/ob livers of tri- and di-acylglycerol species, diacylphosphoglycerols as well as specific reactive ceramide species. Sphingomyelins, the substrate for ceramide synthesis, appeared downregulated in the liver of the ob/ob mice compared to their lean littermates. The increase of acylglycerols should therefore be considered the hallmark leading to the development of the fatty liver observed in the ob/ob mice [42,43].

In order to include the correlation structure of lipidomics data into the analysis and therefore explore possible associations between different lipid molecular species, we applied the partial least squares discriminant analysis (PLS/DA) [44,45] using the SIMPLS algorithm to calculate

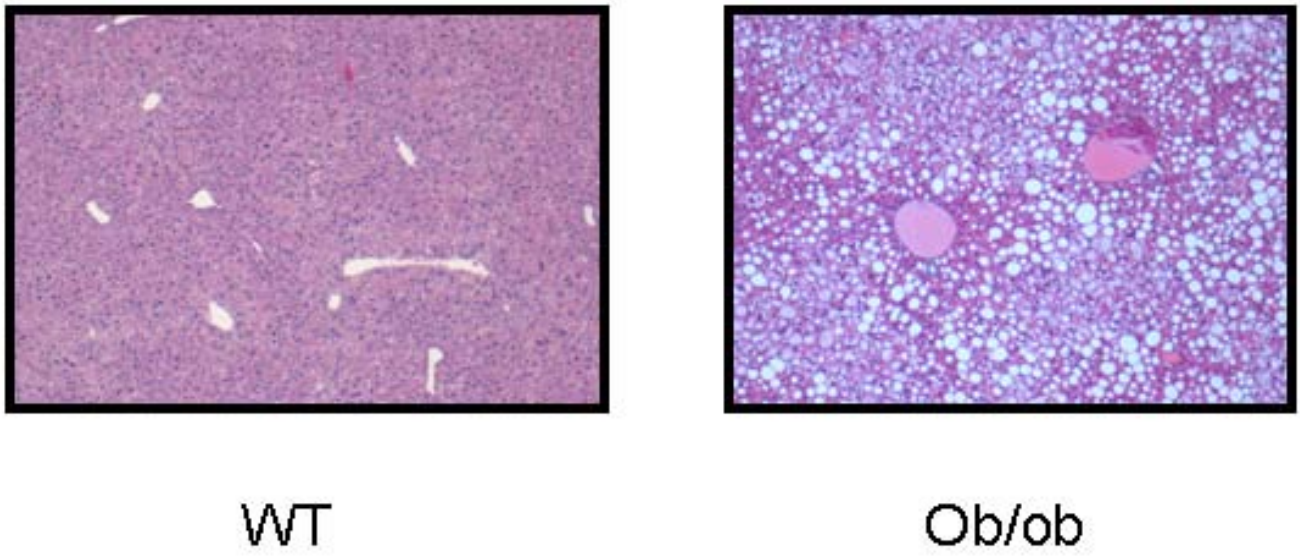


Figure 4
Wild type and ob/ob mice liver cells. Liver tissue images of wild type (WT) and ob/ob mice.

the model [46]. PLS/DA is a common approach to multivariate metabolomics data analysis [47,48]. PLS analysis maximizes the product of variance matrix of measured variables (e.g. lipid profile data) and correlation of measured data with properties of interest (e.g. ob/ob and WT groups). Venetian blinds cross-validation method [49] (with 4 splits) and Q^2 scores were used to optimize the model. Two latent variables were included in the model with the Q^2 = 58%, which can be considered as a significant model. Figure 6A shows the score plot of the PLS/DA model, with as expected clear separation between the genotypes.

The loadings in Figure 6B indicated that the observed separation is largely due to accumulation of acylglycerols in ob/ob mouse livers. Of interest most of the ceramides (including the most abundant Cer(d18:1/18:0) and Cer(d18:1/16:0) species) correlated with the short chain triacylglycerols, suggesting accumulation of reactive ceramide species increase in the liver of the ob/ob mice proportionally to the accumulation of triacylglycerol levels. Curiously, similar correlation between ceramides and triacylglicerols was lost when considering the pool of long chain triacylglycerols. Additionally, we observed the separation of lipid profiles based upon gender basis. The correlation between triacylglycerols and ceramides is particularly interesting since reactive ceramide species are believed to play an important role in development of obesity associated insulin resistance [50]. Therefore our results suggest that measurement of triacylglycerol in liver may be a good indirect indicator of other reactive lipid species pathogenically relevant for the development of insulin resistance.

We also investigated linear associations among lipid species by generating a correlation network. In the network, edges between the nodes representing lipid species are drawn if the Pearson correlation meets the cutoff criterion ($r > 0.75$ and p-value < 0.001). The nodes are colored based on fold change values comparing the mean lipid levels of obese and WT mice. Interestingly, the network corresponding to WT mouse liver sample contains almost double the number of edges (2073) as compared to the number of edges (1055) in the ob/ob mouse liver sample network. Selected clusters of co-regulated lipids corresponding to wild type and ob/ob mouse liver samples are shown in Figure 7. The observed decrease in the number of correlations among the lipid species under ob/ob condition as compared to WT suggests decreased level of co-regulation among lipid species in the ob/ob mouse liver tissue, which can be attributed to ob/ob organ-specific preferential enrichment of subset of lipids. Confirming PLS/DA results, association of ceramides and triacylglycerols is also observed using correlation network analysis.

We then selected two lipid species, TG(54:3) and Cer(d18:1/18:0) from the Figure 7B, and mapped them into the glycerolipid [51] and sphingolipid [52] reference pathways, respectively (Figure 8). While the notation TG(54:3) is redundant as there may be several corresponding lipid molecular species with the same functional group, total number of acyl carbons and double

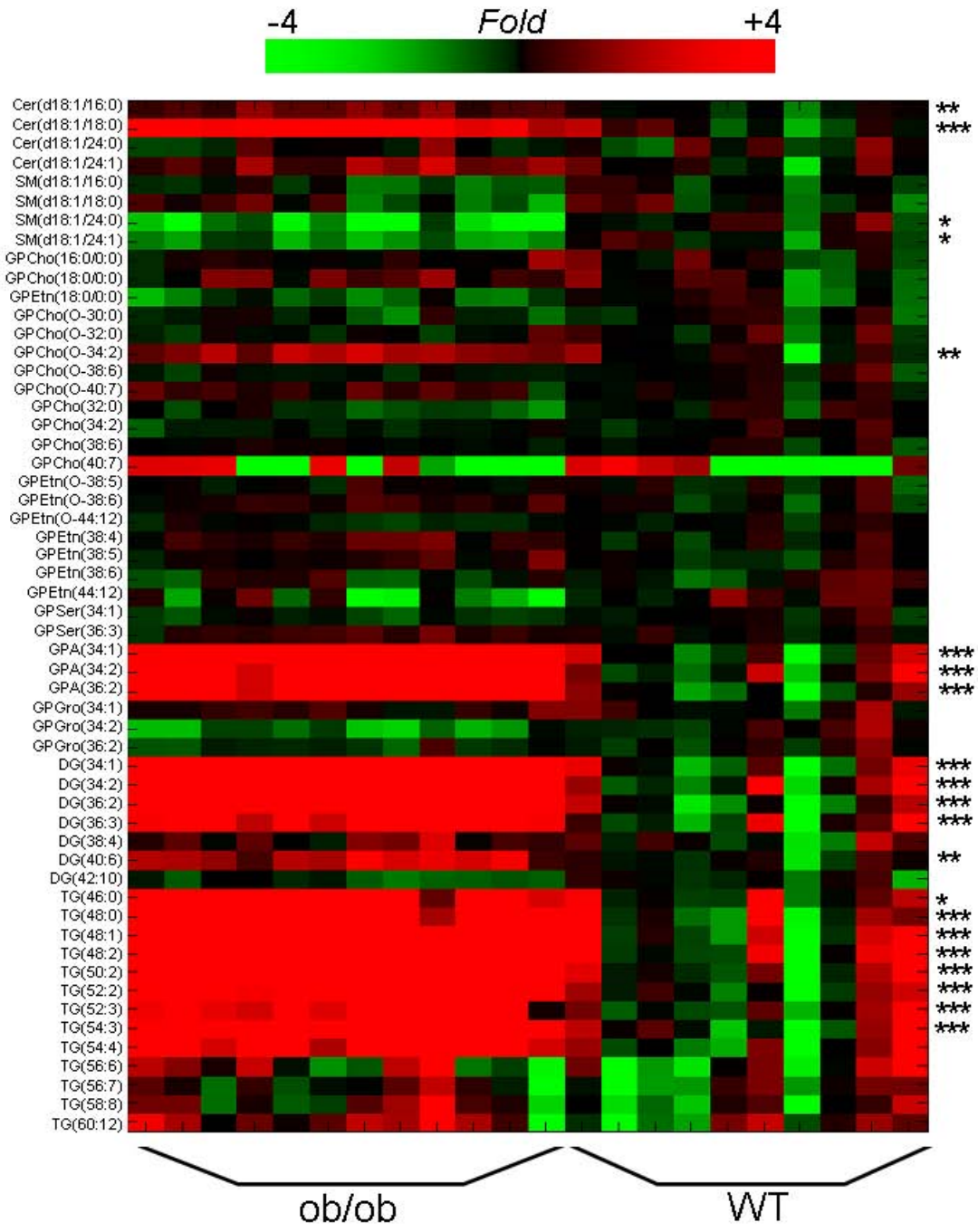


Figure 5

Selected liver lipid profiles from the ob/ob mouse model. Array view of the lipid profiles. The changes are relative to the median intensity of individual molecular species within the Wild Type group. The *p*-values were calculated based on two-sided *t*-test, conservatively adjusted by a Bonfferoni correction for the total number of 192 identified lipids: *($p < 0.05$), **($p < 0.01$), ***($p < 0.001$).

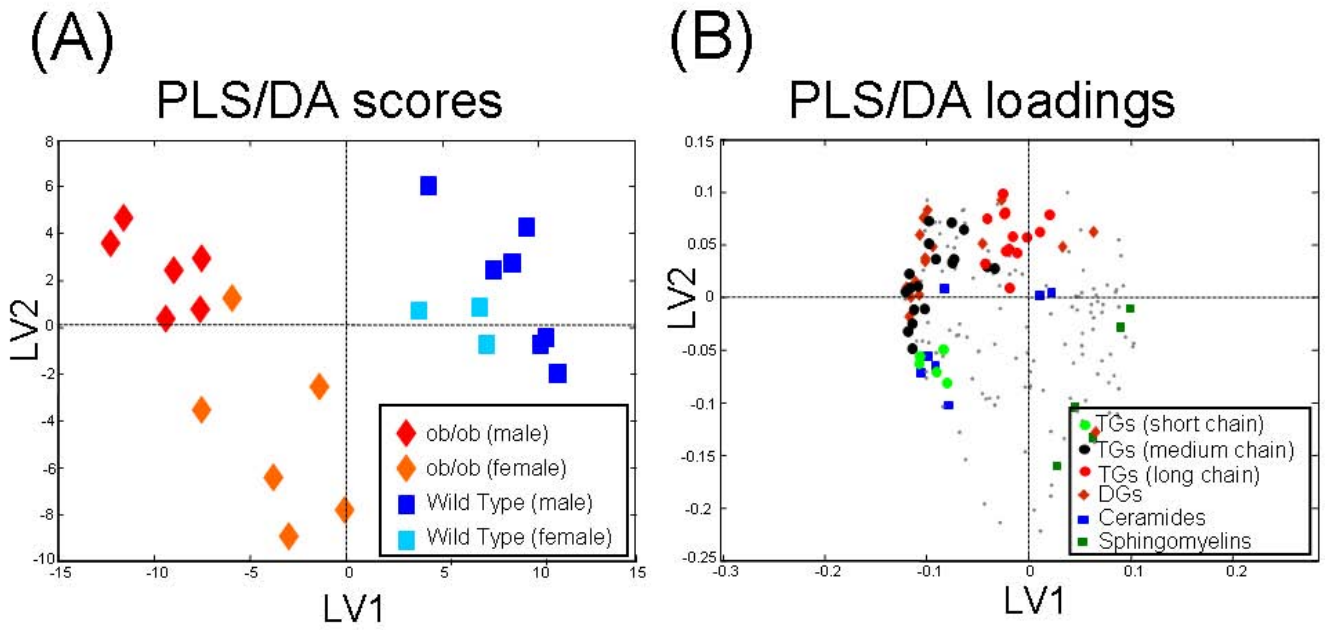


Figure 6
PLS/DA analysis of the ob/ob mouse model. (A) Score plot reveals genotype differences, as well as gender specific differences. (B) Loadings reveal major lipid classes associated with genotype differences.

bonds, we selected one particular instance, TG(18:1/18:1/18:1), for pathway representation. The Figure 8A shows how this particular lipid species is located in the enzymatic system of the glycerolipid pathway. The other pathway, sphingolipid pathway, is instantiated from the co-regulated network with the ceramide lipid species Cer(d18:1/18:0). We utilize for illustration the only publicly available liver ob/ob mouse gene expression data from ChipperDB [53], obtained from 2 month old male mice.

From the sphingolipid pathway map (Figure 8B) two enzymes linked to the ceramide *via* metabolic reactions, one is SGPP1 (Sphingosine-1-phosphate phosphatase 1, UniprotID Q9JI99), the other GALC (galactosylceramidase, UniprotID P54818) were upregulated in ob/ob. SGPP1 is involved in *de novo* ceramide synthesis, while GALC is involved in release of ceramide from glycosphingolipids. Interestingly, sphingomyelin SM(d18:1/18:0) as the known precursor of ceramide via the sphingomyelinase enzymatic action is downregulated, while the sphingomyelinase level is maintained. Therefore, these results indicate that both glycolipids and free fatty acids may contribute as a source of the elevated ceramides observed in the ob/ob fatty liver. The elevated fatty acid flux into the peripheral tissues is a known factor leading to increased ceramide synthesis [50]. In contrast, mobilization of glycosphingolipids for the synthesis of ceramide has not yet been characterized in context of obesity or insulin resistance, although the importance of glycosphingolipids in regulation of insulin sensitivity has been recognized [54]. This is now clearly one area to be investigated further.

Conclusion

Our lipid informatics strategy greatly facilitated interpretation of ob/ob mouse liver lipidomic profiles which resulted in identification of several lipid molecular species. Notable changes in mean lipid levels comparing obese and their normal littermates among the identified lipids included upregulation of tri- and di-acylglycerol species, diacylphosphoglycerols and specific ceramide species, and downregulation of sphingomyelins in ob/ob mice. Correlation network analysis revealed decreased level of co-regulation among lipid species in the ob/ob condition reflecting the specific enrichment of subset of lipids. We observed associations of short and medium chain triacylglycerols and ceramides, both in ob/ob and WT mice, although these species were significantly upregulated in ob/ob mice. The pathway instantiation of specific lipid molecular species in combination to available gene expression data revealed that both glycolipids and free fatty acids are the sources of elevated ceramides in ob/ob fatty liver.

Methods

Database implementation

The lipid data is stored in a native XML database implemented in Tamino XML Server (Software AG). Each com-

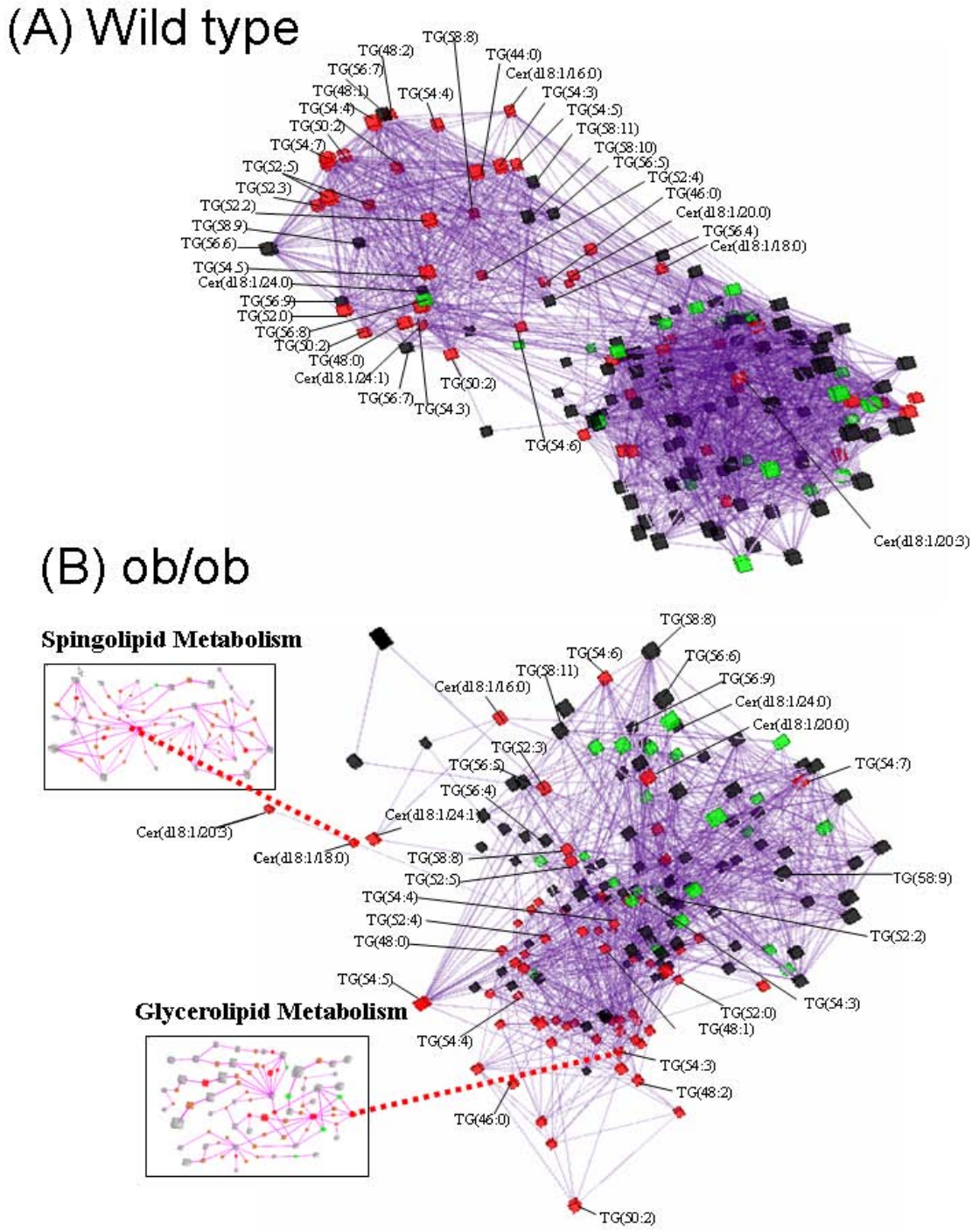


Figure 7

Correlation networks for ob/ob and WT mouse liver lipid profiles. Selected co-regulated cluster of lipid molecular species, including ceramides and acylglycerols. This network is based on Pearson correlation coefficient, r > 0.75 and statistic p-value < 0.001. Colored nodes in red (up regulation), green (down regulation) and black (no change) are based on 1.5-fold change cut off on mean value comparisons for ob/ob *vs*. WT mice. (A) Correlation network for WT mouse liver lipidomic data (B) Correlation network for ob/ob mouse liver lipidomic data.

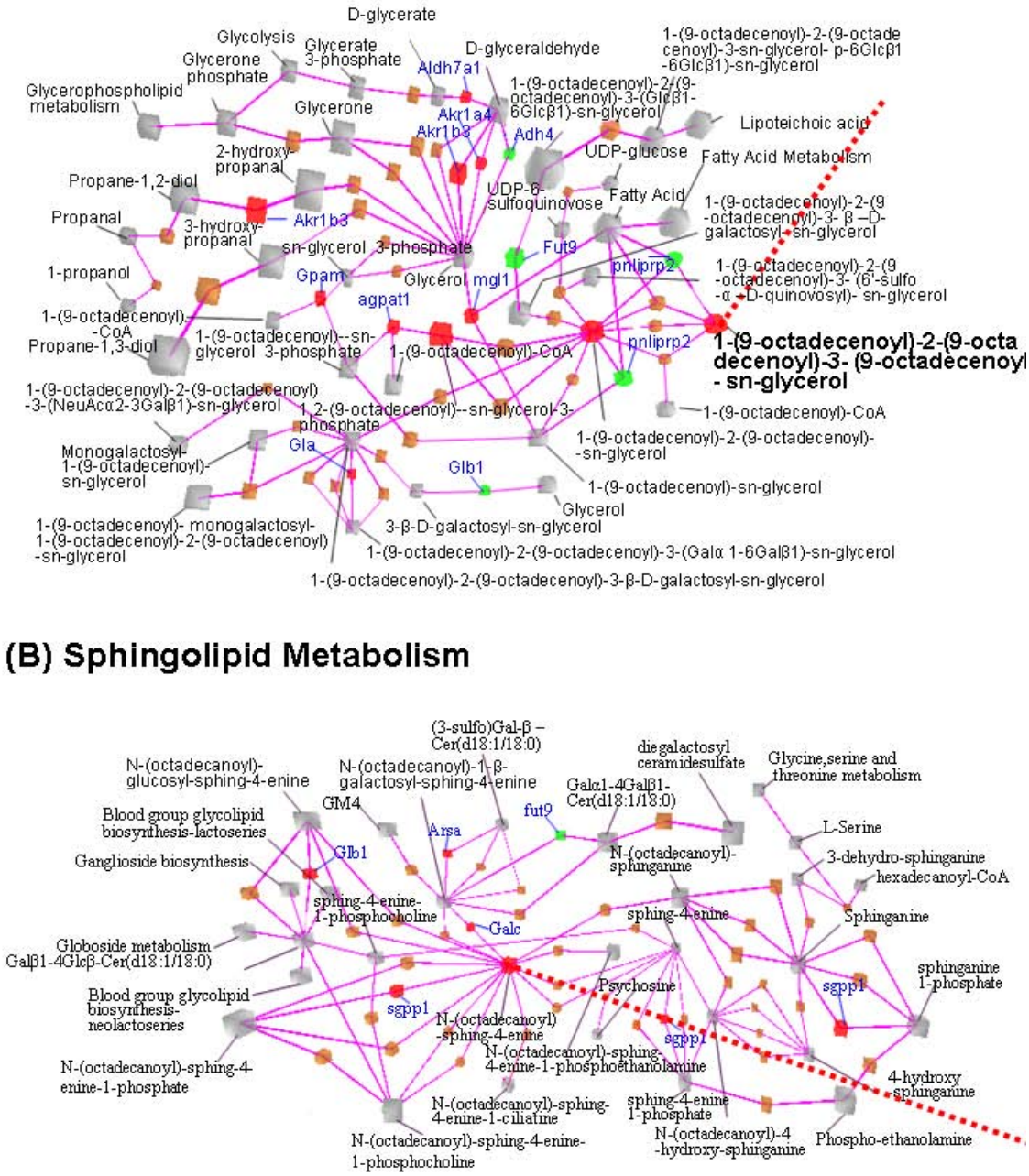


Figure 8

Instantiated pathways for Cer(d18:1/18:0) and TG(18:1/18:1/18:1). Each node represents either lipid metabolite or enzyme or other interconnecting metabolism. Grey color represents metabolites/other metabolism and brown represents enzymes. Up and down regulation of corresponding nodes are denoted by red and green colors, respectively. Enzyme names are shown only if they are differentially regulated (~1.5 fold change). A) Instantiation of TG(18:1/18:1/18:1) as part of glycerolipid metabolism. (B) Instantiation of Cer(d18:1/18:0) as part of sphingolipid metabolism.

pound entry in the database is described by an internal identifier, scoring information, class, canonical SMILES, molecular formula, molecular weight and isotopic distribution. Perl scripts were developed to convert the data into XML documents. Resulting XML documents are loaded using mass-loading tool of the Tamino database.

In the course of implementing the above steps we made use of XMLSPY software (Altova, Inc.) and Tamino Schema Editor Software (Software AG) for the construction and validation of logical and physical schemas, respectively.

Heuristic lipid database scoring

In order to facilitate database searches, an assigned scoring value for each compound in the database is computed from scoring values of seed fatty acid chains from which that compound is formed. Common factors considered while assigning the scoring to seed fatty acid chains are natural abundance of the fatty acid and odd/even number of carbon atoms present in a fatty acid chain. In addition, different type of bonding (e.g. linked via ether bonds) of fatty acids to glycerol back bone carbon gets different score. The lesser the score the more likely the compound is found in the nature. The total score of a lipid is then a product of all fatty acid scores. Random score S of any lipid compound with one or more fatty acid chains whose score variables V_i (at Sn1 position), V_j (at Sn2 position) and V_k(at Sn3 position) is obtained as follows

$$S = \begin{cases} V_i \text{ or } V_j & \text{For compounds with single fatty acid chains (at sn1 or sn2)} \\ V_i \times V_j & \text{For compounds with two fatty acid chains (at sn1 and sn2)} \\ V_i \times V_j \times V_k & \text{For compounds with three fatty acid chains (at sn1, sn2 and sn3)} \end{cases}$$

Animal model background information

Animals were housed at a density of four animals per cage in a temperature-controlled room (20–22*f*C) with 12-h light/dark cycles. Food and water were available *ad libitum*. All animal protocols used in this study were approved by the UK Home Office and the University of Cambridge.

The following genotypes were used for analysis: WT (Lep +/Lep +) and ob/ob (Lep ob/Lep ob). The study included 12 ob/ob (6 male, 6 female) and 10 WT (7 male, 3 female) mice of 16-week age. Genotyping for the point mutation in the ob gene was performed by PCR using standard protocols.

For light microscopy analysis, liver tissues were carefully dissected and fixed in 10% formalin. Tissue was embedded in paraffin and sectioned using a standard microtome (Leica RM2125RT, Leica, UK). Sections were stained with hematoxylin and eosin (H&E) using standard protocols.

Lipidomic analysis

An aliquot of 20 μl of an internal standard mixture, 50 μl of 0.15 M sodium chloride and of chloroform: methanol (2:1) (200 μl) was added to the weighed (20–30 mg) tissue sample. The standard mixture contains the following lipids: GPCho(17:0/17:0) (10 μg/ml), GPEtn(17:0/17:0) (90 μ/ml), GPCho(17:0/0:0) (320 μg/ml), Cer(d18:1/17:0) (90 μg/ml) and TG(17:0/17:0/17:0) (100 μg/ml).

The sample was homogenized and vortexed (2 min for liver or 15 sec for islets)) and after 1 hour for liver or 20 min for islets standing centrifuged at 10000 RPM for 3 min. From the separated lower phase, an aliquot was mixed with 10 μl of a labeled standard mixture (10 μg/ml GPCho(16:0/0:0-D3), GPCho(16:0/16:0-D6) and TG(16:0/16:0/16:0-^{13}C3) and 0.5–1.0 μl injection was used for LC/MS analysis.

Total lipid extracts were analysed on a Waters Q-Tof Premier mass spectrometer combined with an Acquity Ultra Performance LC™ (UPLC). The column, which was kept at 50°C, was an Acquity UPLC™ BEH C18 10 × 50 mm with 1.7 μm particles. The binary solvent system (flow rate 0.200 ml/min) included A. water (1% 1 M NH4Ac, 0.1% HCOOH) and B. LC/MS grade (Rathburn) acetonitrile/isopropanol (5:2, 1% 1 M NH4Ac, 0.1% HCOOH). The gradient started from 65% A/35% B, reached 100% B in 6 min and remained there for the next 7 min. The total run time per sample including a 5 min re-equilibration step was 18 min. The temperature of the sample organizer was set at 10°C.

Mass spectrometry was carried out on Q-Tof Premier (Waters, Inc.) run in ESI+ mode. The data was collected over the mass range of m/z 300–1200 with a scan duration of 0.2 sec. The source temperature was set at 120°C and nitrogen was used as desolvation gas (800 L/h) at 250°C. The voltages of the sampling cone and capillary were 39 V and 3.2 kV, respectively. Reserpine (50 μg/L) was used as the lock spray reference compound (5 μl/min; 10 sec scan frequency).

Lipid identification was performed using tandem mass spectrometry in negative and positive ion mode, as recently described [32].

Authors' contributions

LY developed the lipid informatics methodology, performed data analyses, and drafted the manuscript. MK developed method for processing of UPLC/MS lipidomics data. GMG performed the experiments with ob/ob and WT animals. TSL performed the lipidomics analysis. AVP coordinated the *in vivo* studies and drafted the manuscript. MO initated the study, performed data analyses

and drafted the manuscript. All authors read and approved the final manuscript.

Additional material

Additional File 1

Table of seed fatty acids. The table lists the fatty acids utilized for the lipid scaffold generation.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-1-12-S1.pdf]

Additional File 2

Lipid database (LipidDB) contents. The table lists different lipid classes contained in the database utilized in the paper and their sizes in the database.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-1-12-S2.pdf]

Acknowledgements

This work was supported by Academy of Finland (Grants No 111338 and 207492), Marie Curie International Reintegration Grant from European Community, the Wellcome Trust Integrative Physiology program, and by the EU FP 6 Hepadip integrated program LSHM-CT-2005-018734. We thank C. Ronald Kahn for permission to use the gene expression dataset.

References

1. Vance DE, Vance JE: **Biochemistry of lipids, lipoproteins and membranes.** 4th edition. Edited by: Bernardi G. Amsterdam, The Netherlands, Elsevier B. V.; 2004.
2. Fahy E, Subramaniam S, Brown HA, Glass CK, Merrill AH Jr., Murphy RC, Raetz CRH, Russell DW, Seyama Y, Shaw W, Shimizu T, Spener F, van Meer G, VanNieuwenhze MS, White SH, Witztum JL, Dennis EA: **A comprehensive classification system for lipids.** *J Lipid Res* 2005, **46:**839-862.
3. Wenk MR: **The emerging field of lipidomics.** *Nat Rev Drug Discov* 2005, **4:**594-610.
4. van Meer G: **Cellular lipidomics.** *EMBO J* 2005, **24:**3159-3165.
5. Watson AD: **Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Lipidomics: a global approach to lipid analysis in biological systems.** *J Lipid Res* 2006, **47:**2101-2111.
6. Lagarde M, Geloen A, Record M, Vance D, Spener F: **Lipidomics is emerging.** *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 2003, **1634:**61.
7. **LIPID MAPS** [http://www.lipidmaps.org]
8. Cotter D, Maer A, Guda C, Saunders B, Subramaniam S: **LMPD: LIPID MAPS proteome database.** *Nucl Acids Res* 2006, **34:**D507-510.
9. **LipidBank** [http://www.lipidbank.jp]
10. **CyberLipids** [http://www.cyberlipid.org/]
11. **LIPIDAT** [http://www.lipidat.chemistry.ohio-state.edu/home.stm]
12. Hermansson M, Uphoff A, Kakela R, Somerharju P: **Automated quantitative analysis of complex lipidomes by liquid chromatography/mass spectrometry.** *Anal Chem* 2005, **77:**2166-2175.
13. Houjou T, Yamatani K, Imagawa M, Shimizu T, Taguchi R: **A shotgun tandem mass spectrometric analysis of phospholipids with normal-phase and/or reverse-phase liquid chromatography/electrospray ionization mass spectrometry.** *Rapid Comm Mass Spectrom* 2005, **19:**654-666.
14. Guan XL, He X, Ong WY, Yeo WK, Shui G, Wenk MR: **Non-targeted profiling of lipids during kainate-induced neuronal injury.** *FASEB J* 2006, **20:**1152-1161.
15. Bijlsma S, Bobeldijk I, Verheij ER, Ramaker R, Kochhar S, Macdonald IA, vanOmmen B, Smilde AK: **Large-scale human metabolomics studies: A strategy for data (pre-) processing and validation.** *Anal Chem* 2006, **78:**567-574.
16. Ekroos K, Chernushevich IV, Simons K, Shevchenko A: **Quantitative profiling of phospholipids by multiple precursor ion scanning on a hybrid quadrupole time-of-flight mass spectrometer.** *Anal Chem* 2002, **74:**941 -9949.
17. Han X, Gross RW: **Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: a bridge to lipidomics.** *J Lipid Res* 2003, **44:**1071-1079.
18. Schwudke D, Oegema J, Burton L, Entchev E, Hannich JT, Ejsing CS, Kurzchalia T, Shevchenko A: **Lipid profiling by multiple precursor and neutral loss scanning driven by the data-dependent acquisition.** *Anal Chem* 2006, **78:**585-595.
19. McAnoy AM, Wu CC, Murphy RC: **Direct qualitative analysis of triacylglycerols by electrospray mass spectrometry using a linear ion trap.** *J Am Soc Mass Spectrom* 2005, **16:**1498-1509.
20. Lu Y, Hong S, Tjonahen E, Serhan CN: **Mediator-lipidomics: databases and search algorithms for PUFA-derived mediators.** *J Lipid Res* 2005, **46:**790-802.
21. Katajamaa M, Oresic M: **Processing methods for differential analysis of LC/MS profile data.** *BMC Bioinformatics* 2005, **6:**179.
22. Katajamaa M, Miettinen J, Oresic M: **MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data.** *Bioinformatics* 2006, **22:**634-636.
23. Stolt R, Torgrip RJO, Lindberg J, Csenki L, Kolmert J, Schuppe-Koistinen I, Jacobsson SP: **Second-order peak detection for multicomponent high-resolution LC/MS data.** *Anal Chem* 2006, **78:**975-983.
24. Prakash A, Mallick P, Whiteaker J, Zhang H, Paulovich A, Flory M, Lee H, Aebersold R, Schwikowski B: **Signal maps for mass spectrometry-based comparative proteomics.** *Mol Cell Proteomics* 2006, **5:**423-432.
25. Smilde AK, vanderWerf MJ, Bijlsma S, vanderWerff-vanderVat BJC, Jellema RH: **Fusion of mass spectrometry-based metabolomics data.** *Anal Chem* 2005, **77:**6729-6736.
26. Oresic M, Clish CB, Davidov EJ, Verheij E, Vogels JTWE, Havekes LM, Neumann E, Adourian A, Naylor S, Greef J, Plasterer T: **Phenotype characterization using integrated gene transcript, protein and metabolite profiling.** *Appl Bioinformatics* 2004, **3:**205-217.
27. Medina-Gomez G, Virtue S, Lelliott C, Boiani R, Campbell M, Christodoulides C, Perrin C, Jimenez-Linan M, Blount M, Dixon J, Zahn D, Thresher RR, Aparicio S, Carlton M, Colledge WH, Kettunen MI, Seppanen-Laakso T, Sethi JK, O'Rahilly S, Brindle K, Cinti S, Oresic M, Burcelin R, Vidal-Puig A: **The link between nutritional status and insulin sensitivity is dependent on the adipocyte-specific Peroxisome Proliferator-Activated Receptor-{gamma}2 isoform.** *Diabetes* 2005, **54:**1706-1716.
28. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: **The KEGG resource for deciphering the genome.** *Nucl Acids Res* 2004, **32:**D277-280.
29. Weininger D: **Hanbbook of chemoinformatics - from data to knowledge.** *Volume 1*. Edited by: Gasteiger J. , Wiley-VCH Verlag GmBH & Co.KGaA,Weinheim; 2003.
30. Weininger D: **SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules.** *J Chem Inf Comput Sci* 1988, **28:**31-36.
31. Lide DR: **CRC Handbook of chemistry and physics.** 85th edition. , CRC Press; 2004.
32. Laaksonen R, Katajamaa M, Päivä H, Sysi-Aho M, Saarinen L, Junni P, Lütjohann D, Smet J, Coster RV, Seppänen-Laakso T, Lehtimäki T, Soini J, Oresic M: **A systems biology strategy reveals biological pathways and plasma biomarker candidates for potentially toxic statin induced changes in muscle.** *PLoS ONE* 2006, **1:**e97.
33. **Isotope Pattern Calculator** [http://isotopatcalc.sourceforge.net]
34. Camacho D, de la Fuente A, Mendes P: **The origin of correlations in metabolomics data.** *Metabolomics* 2005, **1:**53-63.
35. Kose F, Weckwerth W, Linke T, Fiehn O: **Visualizing plant metabolomic correlation network using clique-metabolite matrices.** *Bioinformatics* 2001, **17:**1198-1208.
36. Rischer H, Oresic M, Seppanen-Laakso T, Katajamaa M, Lammertyn F, Ardiles-Diaz W, Van Montagu MCE, Inze D, Oksman-Caldentey KM, Goossens A: **Gene-to-metabolite networks for terpenoid**

indole alkaloid biosynthesis in Catharanthus roseus cells. *PNAS* 2006, **103:**5614-5619.
37. Steuer R, Kurths J, Fiehn O, Weckwerth W: **Observing and interpreting correlations in metabolomic networks.** *Bioinformatics* 2003, **19:**1019-1026.
38. Gopalacharyulu PV, Lindfors E, Bounsaythip C, Kivioja T, Yetukuri L, Hollmen J, Oresic M: **Data integration and visualization system for enabling conceptual biology.** *Bioinformatics* 2005, **21:**i177-185.
39. Adiels M, Packard C, Caslake MJ, Stewart P, Soro A, Westerbacka J, Wennberg B, Olofsson SO, Taskinen MR, Boren J: **A new combined multicompartmental model for apolipoprotein B-100 and triglyceride metabolism in VLDL subfractions.** *J Lipid Res* 2005, **46:**58-67.
40. Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM: **Positional cloning of the mouse obese gene and its human homologue.** *Nature* 1994, **372:**425-432.
41. Browning JD, Horton JD: **Molecular mediators of hepatic steatosis and liver injury.** *J Clin Invest* 2004, **114:**147-152.
42. Yang SQ, Lin HZ, Lane MD, Clemens M, Diehl AM: **Obesity increases sensitivity to endotoxin liver injury: Implications for the pathogenesis of steatohepatitis.** *PNAS* 1997, **94:**2557-2562.
43. Lin HZ, Yang SQ, Chuckaree C, Kuhajda F, Ronnet G, Diehl AM: **Metformin reverses fatty liver disease in obese, leptin-deficient mice.** *Nat Med* 2000, **6:**998-1003.
44. Geladi P, Kowalski BR: **Partial least-squares regression: a tutorial.** *Anal Chim Acta* 1986, **185:**1-17.
45. Barker M, Rayens W: **Partial least squares for discrimination.** *J Chemometrics* 2003, **17:**166-173.
46. de Jong S: **SIMPLS: An alternative approach to partial least squares regression.** *Chemometr Intell Lab Syst* 1993, **18:**251-263.
47. Pears MR, Cooper JD, Mitchison HM, Mortishire-Smith RJ, Pearce DA, Griffin JL: **High resolution 1H NMR-based metabolomics indicates a neurotransmitter cycling deficit in cerebral tissue from a mouse model of Batten Disease.** *J Biol Chem* 2005, **280:**42508-42514.
48. Brindle JT, Antti H, Holmes E, Tranter G, Nicholson JK, Bethell HWL, Clarke S, Schofield PM, McKilligin E, Mosedale DE, Grainger DJ: **Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics.** *Nat Med* 2002, **8:**1439-1445.
49. Wise BM, Gallagher NB, Bro R, Shaver JM, Windig W, Koch JS: **PLS Toolbox 3.5 for use with Matlab.** Manson, WA, Eigenvector Research Inc.; 2005.
50. Summers SA: **Ceramides in insulin resistance and lipotoxicity.** *Prog Lipid Res* 2006, **45:**42-72.
51. **KEGG Glycerolipid Metabolism** [http://www.genome.jp/kegg/pathway/map/map00561.html]
52. **KEGG Sphingolipid Metabolism** [http://www.genome.jp/kegg/pathway/map/map00600.html]
53. **ChipperDB: Diabetes Genome Anatomy Project** [http://www.diabetesgenome.org/chipperdb/expt.cgi?id=65]
54. Yamashita T, Hashiramoto A, Haluzik M, Mizukami H, Beck S, Norton A, Kono M, Tsuji S, Daniotti JL, Werth N, Sandhoff R, Sandhoff K, Proia RL: **Enhanced insulin sensitivity in mice lacking ganglioside GM3.** *PNAS* 2003, **100:**3445-3449.

PUBLICATION III

Composition and lipid spatial distribution of High Density Lipoprotein particles in subjects with low and high HDL-cholesterol

In: Journal of Lipid Research 2010.
In press.

Composition and lipid spatial distribution of High Density Lipoprotein particles in subjects with low and high HDL-cholesterol

Laxman Yetukuri,[1] Sanni Söderlund,[2] Artturi Koivuniemi,[3] Tuulikki Seppänen-Laakso,[1], Perttu S. Niemelä,[1] Marja Hyvönen,[4] Marja-Riitta Taskinen,[2] Ilpo Vattulainen,[3,5,6] Matti Jauhiainen,[7,8] and Matej Orešič[1,8,*]

[1]VTT Technical Research Centre of Finland, Espoo, Finland, [2]Division of Cardiology, Department of Medicine, University of Helsinki, Helsinki, Finland, [3]Department of Physics, Tampere University of Technology, Tampere, Finland, [4]Department of Physics, University of Oulu, Finland, [5]Department of Physics, Aalto University School of Science and Engineering, Espoo, Finland, [6]MEMPHYS – Center for Biomembrane Physics, University of Southern Denmark, Odense, Denmark, [7]National Institute for Health and Welfare, Helsinki, Finland, [8]Institute for Molecular Medicine Finland FIMM, Helsinki, Finland

*Correspondence to: Matej Orešič, VTT Technical Research Centre of Finland, P.O. Box 1000, Espoo, FI-02044 VTT, Finland. Tel.: +358 20 722 4491; Fax: +358 20 722 7071; Email: matej.oresic@vtt.fi.

Abbreviations: ChoE – cholesteryl ester; FCho – free cholesterol; FDR – false discovery rate ; HDL-C – high-density lipoprotein cholesterol; LDL-C – low-density lipoprotein cholesterol; lysoPC – lysophosphatidylcholine; MS – mass spectrometry; PC – phosphatidylcholine; PE – phosphatidylethanolamine; PLS/DA – partial least square discriminant analysis; RCT – reverse cholesterol transport; SM – sphingomyelin; TG – triacylglycerol; UPLC^{TM} – Ultra Performance Liquid Chromatography**.**

Low level of high density lipoprotein cholesterol (HDL-C) is a powerful risk factor for cardiovascular disease. However, despite the reported key role of apolipoproteins specifically apoA-I in HDL metabolism, lipid molecular composition of HDL particles in subjects with high and low HDL-C levels is currently unknown. Here lipidomics was used to study HDL derived from well characterized high and low HDL-C subjects. Low HDL-C subjects had elevated triacylglycerols and diminished lysophosphatidylcholines and sphingomyelins. Using information about the lipid composition of HDL particles in these two groups, we reconstituted HDL particles *in silico* by performing large scale molecular dynamics simulations. In addition to confirming the measured change in particle size, we found that the changes in lipid composition also induced specific spatial distributions of lipids within the HDL particles, including higher amount of triacylglycerols at the surface of HDL particles in low HDL-C subjects. Our findings have important implications for understanding HDL metabolism and function. For the first time we demonstrate the power of combining molecular profiling of lipoproteins with dynamic modeling of lipoprotein structure.

Keywords: High Density Lipoprotein / lipidomics / lipid metabolism / molecular dynamics

INTRODUCTION

High-density lipoprotein (HDL) is one of the five major lipoproteins (chylomicrons, VLDL, IDL, LDL, HDL). HDL is the smallest and densest of the lipoproteins because it contains the highest proportion of protein. Low level of HDL cholesterol (HDL-C) is a powerful risk factor for cardiovascular disease (1-3). Accumulating evidence suggests, however, that HDL-C alone may not be an adequate marker of atheroprotection. HDLs are compositionally and functionally diverse lipoprotein particles, which needs to be taken into account in the evaluation of cardiovascular risk (4, 5). Recent study of HDL proteome revealed many important changes in the protein composition of HDL in cardiovascular patients, without changes in serum HDL-C (6).

Detailed characterization of lipoprotein fractions and changes in their molecular lipid sprofile may help identifying novel biomarkers in lipid metabolism (7). Main lipid constituents of HDL particles include glycerophospholipids, cholesteryl esters (ChoE), sphingomyelins (SM) and triacylglycerols (TG). Lysophosphatidylcholines (lysoPC) are known to be associated with pro-atherogenic conditions (8). Enrichment of HDL phospholipids such as phosphatidylcholines (PC) and SM improves the net efflux of cholesterol from scavenger receptor-BI expressing cells (9) and phospholipid composition may have a major impact in the process of reverse cholesterol transport (RCT) (10). We also demonstrated using HDL derived either from low or high HDL-C subjects that cholesterol efflux from human THP-1 macrophages correlated with phospholipids, particle size and particle mass of HDL (11). This is consistent with earlier efflux studies demonstrating that the phospholipid content of HDL is an important determinant of cholesterol egress (12, 13). Additionally, the activity of HDL in the first step of RCT is affected by fatty acyl chain length of

the phospholipids (14). Together, the information on detailed lipid molecular composition of HDL may provide better insights into the mechanisms behind the anti-atherogenic role of HDL particles.

Conventional lipoprotein analyses have relied on analyses of total protein, phospholipids, free cholesterol, ChoE and TG content (15). However, molecular level concentrations may provide more precise markers of specific metabolic phenotypes than total lipid class concentrations (16). Recent advances in mass spectrometry (MS) based analytical platforms and bioinformatics approaches for managing large volumes of the data have made it possible to study lipid species at the molecular level (17, 18). The lipidomics platform based on Ultra Performance Liquid Chromatography Mass Spectrometry (UPLC^{TM}/MS) was recently utilized to characterize molecular lipids including triacylglycerols, glycerophospholipids, sphingomyelins, cholesteryl esters and ceramides in different lipoprotein fractions (16).

Lipid molecular composition of HDL particles in subjects with high and low HDL-cholesterol levels has not been studied systematically. Here we combine clinical cohort study and global lipidomics with molecular simulations of HDL particles. We apply UPLC/MS-based lipidomics to study HDL fractions from well characterized high and low HDL-C subjects from a large Finnish population cohort and identify many specific changes in HDL lipidomes between subjects with high and low HDL-C. Using information about the lipid composition of HDL particles in these two groups, we reconstitute HDL particles *in silico* by performing large scale molecular dynamics simulations. In addition to confirming the measured change in particle size, we show that HDL particles derived from high HDL-C subjects have surprisingly different spatial distribution of triacylglycerols.

MATERIALS AND METHODS

Study subjects

The study comprised 47 subjects: 24 low-HDL subjects and 23 high-HDL subjects who were participants of the Health 2000 Health Examination Survey (19). The subjects represented the extreme ends of HDL-C levels ($\leq 10^{th}$ and $\geq 90^{th}$ percentiles) and the HDL-C limits were as follows: for low-HDL-C men $\leq$ 1.03 mmol/l, low-HDL-C women $\leq$ 1.23 mmol/l, high-HDL-C men $\geq$ 1.79 mmol/l, and high-HDL-C women $\geq$ 2.24 mmol/l. Subjects with diabetes, alcohol abuse, or malignancy were excluded. Alcohol abuse was defined as >160 grams of alcohol / week for women and >310 grams of alcohol / week for men. In addition, subjects using systemic estrogen, corticosteroid therapy, statins or other drugs affecting HDL metabolism were excluded. Each study subject gave a written informed consent before participating in the study. The samples were collected in accordance with the Helsinki declaration and the ethics committees of the participating centers approved the study design.

Lipoprotein separation and characterization

HDL for the lipidomic analysis was separated from plasma samples by ultracentrifugation (20), HDL subspecies distribution and HDL mean particle size were determined with native gradient gel electrophoresis (21) with minor modifications as previously described (11). The molecular size intervals for HDL subspecies 2b, 2a, 3a, 3b, and 3c were used according to Blanche *et al*. (21), and for each subspecies, the relative area under the densitometric scan is reported. HDL mean particle size was calculated by multiplying the mean size of each HDL subclass by its relative area under the densitometric scan (22). LDL peak particle size was measured with gradient gel electrophoresis as previously described in detail (23).

Biochemical analyses

Venous blood samples were drawn after an over-night fast. Serum and EDTA plasma samples were stored at -70°C before analysis. Serum total cholesterol (TC), TG, and HDL-C were measured with Olympus AU400 clinical chemistry analyzer (Olympus, Hamburg, Germany) by fully enzymatic methods (Olympus kits OSR 6116 and 6133 for TC and TG, respectively, and Roche Diagnostics kit 3030024 for HDL-C (Roche Diagnostics GmbH, Mannheim, Germany)). LDL-C was calculated using the Friedewald formula.(24) Concentrations of apoA-I and apolipoprotein B (apoB) were measured with Olympus AU400 analyser by immunoturbidometric methods (kits 64265 and 67249 from Orion Diagnostica, Espoo, Finland). Serum apolipoprotein A-II (apoA-II) was measured with Cobas Mira analyser (Hoffman-La Roche, Basel, Switzerland) immunoturbidometrically (Wako Chemicals GmbH, Neuss, Germany, and own polyclonal antibody produced in rabbits against purified human apoA-II). Serum apoE concentration was quantitated by ELISA (25). Plasma glucose was measured by the glucose dehydrogenase method (Merck Diagnostica, Darmstadt, Germany). Plasma insulin was measured by radioimmunoassay (Pharmacia AB, Uppsala, Sweden).

Phospholipid transfer protein (PLTP) activity was measured using the radiometric assay as previously described (26), with minor modifications (27). PLTP concentration was measured with ELISA (28). Cholesterol ester transfer protein (CETP) activity was measured with radiometric assay as described by Groener *et al*. (29). Paraoxonase activity was measured with spectrophotometry (30). Nitrotyrosine concentration was measured with ELISA (HyCult biotechnology kit HK 501 Uden, The Netherlands). The measurements of IL-6, TNF-α and CRP were done using the Immunochemiluminometric assay (Immulite, DPC, Siemens Healthcare Diagnostics, USA).

Blood pressure values are the mean values from three consecutive measurements carried out in 1-2 minutes intervals. Data on alcohol consumption was collected from questionnaires filled in by the study subjects.

Lipidomic analysis

An internal standard mixture containing 10 lipid compounds was added to each sample (20 µl of total HDL fraction). Lipids were extracted with chloroform / methanol (2:1, v/v, 100µl) solvent. The samples were centrifuged (10000 rpm, 3 min) and 60µl of the lower lipid extract was taken into an HPLC vial insert and another standard mixture containing 3 labelled lipid compounds was added. The internal standards include PC(17:0/0:0), PC(17:0/17:0), PE(17:0/17:0), PG(17:0/17:0)[rac], Cer(d18:1/17:0), PS(17:0/17:0), PA(17:0/17:0) and D-*erythro*-Sphingosine-1-Phosphate (C17 Base) from Avanti Polar Lipids (Alabaster, AL) and MG(17:0/0:0/0:0)[rac], DG(17:0/17:0/0:0)[rac] and TG(17:0/17:0/17:0) from Larodan Fine Chemicals (Malmö, Sweden). The labeled standards include PC(16:0/0:0-D3), PC(16:0/16:0-D6) and TG(16:0/16:0/16:0-^{13}C3 from Larodan Fine Chemicals (Malmö, Sweden).

Lipid extracts (2 µl injections) were analysed on a Waters Q-Tof Premier mass spectrometer combined with an Acquity Ultra Performance LC™ (UPLC) (Waters Inc., Milford, MA). The column was an Acquity UPLC^{TM} BEH C18 10 × 50 mm with 1.7 µm particles and the gradient solvent system included water (1% 1M NH_4Ac, 0.1% HCOOH) and LC/MS grade (Rathburn) acetonitrile/ isopropanol (5:2, 1% 1M NH_4Ac, 0.1% HCOOH). The total run time including a 5 min re-equilibration step was 18 min. The flow rate was 0.200 ml/min. The data were collected at mass range of m/z 300-1200 with a scan duration of 0.2 sec in ESI+ mode.

Statistical methods

Statistical analyses were performed using a freely available R language (http://www.r-project.org/). False discovery rate (FDR) *q*-values were computed using statistical methods from R package 'qvalue'. Correlation analysis was performed using 'gplots' library from R package.

Chemometric modeling of data

Supervised model was built for clustering and discrimination using partial least squares discriminant analysis (PLS/DA) (31, 32). The PLS/DA model attempts to get the latent variables by maximizing the covariance between measured data (x) (*e.g.*, lipid profile data) and response variables of interest (y) (*e.g.* high HDL-C and low HDL-C groups). The model was built by scaling x data to unit variance and zero mean and y data to zero mean. The random subsets cross validation method (33) and Q^2 scores were used to optimize the models. The VIP (variable importance in the projection) values(34) were computed to identify most important lipid species contributing to separation of low- and high- HDL-C groups in the PLS/DA model. PLS/DA model was built using Matlab, version 7.0 (Mathworks, Natick, MA) and PLS Toolbox, version 4.0, of the Matlab package (Eigenvector Research, Wenatchee, WA).

Construction of simulation systems and simulation parameters

In the case of apoA-I, we first build an all-atom model for apoA-I based on the previous data after which we coarse grained the structure. The relative conformation of apoA-Is was similar to the belt-like structure of Borhani *et al.* who produced the solution X-ray structure of truncated apoA-I (residues 44-243) (35). Here, we also followed the approach of Segrest and coworkers who arranged atomistic apoA-Is around the lipid moiety in belt-like fashion (36) so that the

hydrophobic sides of amphiphilic alpha-helices were pointing towards the lipid moiety. This structure was further used to build a model for the full-length apoA-I by adding the previously absent N-terminal part (residues 1-43) of the apoA-I to our model. The conformation of this segment was determined to be alpha-helical as suggested by the previous NMR data (37). The hydrophobic side of this helix was also orientated towards the lipid moiety. We directly coarse grained the atomistic structure and placed the apoA-Is around the previously simulated lipid droplet. The secondary structure of apoA-I was enforced almost completely to the alpha-helical conformation, only the N- and C-terminal ends were modeled as random coils. This approach has been used previously in the work of Catte and coworkers with truncated apoA-I (38). Thus, the flexibility of alpha-helical segments arises from the three-body angle and four-body dihedral potentials of backbone beads that were produced by Monticelli *et al.* in order to extend the MARTINI force field to proteins (39). The radius of the ring formed by apoA-Is was ~12.5 nm and the average distance between apoA-Is was 2.5-3.0 nm.

Next, in order to produce starting structures for the low HDL-C, normal HDL-C and high HDL-C systems the lipid composition of the previously simulated lipid droplet was tuned accordingly to Table 2. All lipid molecules, expect those that were added to the starting system, were inside the ring formed by apoA-Is. Next all three systems were energy minimized by the steepest descent algorithm and 10 ns vacuum simulations were carried out in order to get lipids to the one unified phase meaning that the added lipids diffused to the main lipid particle. Systems were solvated and 18 water beads were changed to Na+ beads in order to neutralize the charges in the systems. Production simulations lasted for 2 μs, which corresponds to 8 μs of effective time since the MARTINI model speeds up the dynamics by an approximate factor of four (39). System sizes

were approximately 30000-35000 water beads and 4300-5000 lipid and protein beads in total. The box size in each simulation was 17 x 17 x 17 nm.

Simulations were performed by the GROMACS simulation package (v. 4.0) (40) and the standard Martini lipid force field was used for PC (PC(16:0/18:1)), lysoPC (PC(16:0/0:0)), FCho and SM (SM(d18:1/16:0)) molecules (41). Protein part was modeled using the protein extension of the Martini description (39), TG (TG(18:1/18:1/18:1)) and ChoE (ChoE(18:1)) force fields were constructed based on the philosophy of Martini force field, the exact parameters and validation are available elsewhere (38). In all simulations temperature was set to 310 K and pressure to 1 bar. Berendsen temperature and pressure coupling algorithms (42) were utilized with coupling constants of 0.4 ps and 2.0 ps, respectively. All lipid classes, apoA-Is, water and ions were separately coupled to heat bath. Electrostatic and Lennard-Jones interactions were calculated using the shift type potentials with cut-off lengths of 1.2 nm and the potentials were shifted to zero starting at 0.0 and 0.9 nm, respectively. Time step was set to 0.025 ps. In each case, the first 4 μs were treated as equilibration simulation and the last 4 μs was used in the analysis. Gromacs analysis programs g_rdf, g_mindist, g_gyrate and g_rmsf were used in the analysis.

RESULTS

Characteristics of the study subjects

Clinical and biochemical characteristics of low and high HDL-C subjects are shown in Table 1. As expected the high HDL-C subjects had higher fasting serum HDL-C ($P < 0.001$) and total cholesterol ($P < 0.001$) concentrations as compared to low HDL-C subjects. However, fasting serum LDL-C concentrations between the groups did not differ significantly. High HDL-C subjects were leaner as compared to low HDL-C subjects, with lower body mass index ($P <$

0.001), lower triglycerides ($P < 0.001$) and insulin ($P = 0.0015$), as well as lower TNF-α ($P = 0.038$) and hs-CRP ($P = 0.012$) levels in plasma. The two groups had similar PLTP mass and CETP activity levels (Table 1). As expected, high HDL-C subjects had higher apoA-I ($P < 0.001$) and lower apoB ($P < 0.001$) values than low HDL-C subjects.

Lipidomic profiling of HDL fractions

To characterize the HDL-associated lipids at the molecular level, the established lipidomics platform using UPLC/MS was applied as described previously (43). A total of 307 molecular lipids across 12 functional classes were detected. The partial least squares discriminant analysis (PLS/DA) revealed that the HDL lipidomic profiles in low HDL-C and high HDL-C subjects are clearly different (Fig. 1A). As shown by clustering of the top ranked discriminant lipids in Fig. 1B, the HDL compositional changes in high HDL-C subjects as compared to low HDL-C subjects were dominated by elevated lysoPCs, SMs, ChoEs and diminished TGs. No within-class trend was observed between the low and high HDL-C groups for PCs. Box plots for the most abundant lipids from lysoPC, SM, ChoE, ethanolamine plasmalogen (PEp) and TG classes are shown in Fig. 2. The most significant identified lipid species with their fold changes and False Discovery Rate (FDR) *q*-values are shown in Table S1. Full lipidomics results for 307 detected lipid peaks are shown in Table S2. The observed highly significant differences between the HDL lipidomic profiles of the high and low HDL-C groups were not attributed to gender (Fig. S1) or TG fatty acid composition (Fig. S2), despite the observed gender differences in the HDL subspecies distribution in the low HDL-C group (Table S3).

Although not being among the most abundant HDL phospholipids, several of the ether lipids were found among the most significantly different in high and low HDL groups (Fig. 1 and Table S2), including the ethanolamine plasmalogens PE(P-16:0/18:2), PE(P-16:0/20:4) and PE(P-18:1/20:4) (Table S2). The ethanolamine plasmalogens as well as PC ether lipids containing high number of double bonds were diminished in the high HDL group, while the more saturated ether PCs were elevated as compared to the levels in the low HDL group.

The top ranked lipids derived from PLS/DA (Fig. 1A) were correlated with the selected clinical variables (Fig. S3). The concentrations of TGs were negatively while the SMs and lysoPCs were positively correlated with the HDL-C concentration. We then investigated associations of HDL-C with the top-ranked lipids from different lipid classes in low and high HDL-C subjects separately (Figs. S4 and S5). Positive correlations between SM(d18:1/16:0) and HDL-C were similar in low HDL-C subjects (r = 0.71, P = 0.0001) and high HDL-C subjects (r = 0.71, P = 0.0001). Interestingly, positive correlation identified between lysoPC(18:0) and HDL-C in low HDL-C subjects (r = 0.54, P = 0.006) was absent in the high HDL-C group (r = -0.06, P = 0.78). We observed no correlation of TG(16:0/18:1/20:1) with HDL-C in low HDL-C subjects (r = 0.06, P = 0.75) whereas there was a negative correlation in high HDL-C subjects (r = -0.43, P = 0.04). No significant correlations were found between ChoE(18:1) and HDL-C, neither in low HDL-C subjects (r = 0.33, P = 0.12) nor in high HDL-C subjects (r = 0.02, P = 0.92). Correlation analysis was also performed for the HDL particle size *vs*. the selected lipids (Figs. S4 and S5). HDL particle size correlated positively with lysoPC (18:0) (r = 0.50, P = 0.0003) and SM(d18:1/16:0) (r = 0.58, P = 0.0003), and negatively with TG(16:0/18.1/20:1) (r = -0.51, P < 0.0001). No such significant correlations were observed when analyzing the low and high HDL-C groups separately.

Simulations of HDL particles

Our findings imply that the lipid composition of HDL particles as derived from subjects with high and low HDL cholesterol is very different. In order to gain further insight how the changes in the lipid composition of HDL may influence HDL structure and function, we performed coarse grained simulations of HDL particles, with their molecular composition derived from within-group mean selected lipid concentrations as obtained by lipidomics. Three coarse grained molecular dynamics simulations with different lipid compositions were carried out. In addition to the "High HDL-C" and "Low HDL-C" particles, we also simulated the "Normal HDL-C" as an intermediate between the two. The numbers of lipids used in each simulation are listed in Table 2. The lipid species concentrations obtained in global lipidomic profiling were expressed relative to the moles of apoA-I content. If lipid species in a certain class were contributing to separation of low and high HDL-C groups as per the PLS/DA model, the averaged concentration of all those species in that class was used to find estimate of fold changes between low and high HDL-C groups. Alternatively, for the lipid species such as cholesteryl esters, which did not contribute much in the separation of two HDL-C groups as per the model, the averaged concentration levels of all identified species were used to find the fold change estimate between the low and high HDL-C groups. Thus obtained fold change values between low and high HDL-C groups were used to estimate the number of species per HDL particle in low and high HDL-C particles by comparing the numbers in the normal-HDL-C compositions. The lipid composition for the normal-HDL model was chosen based on the previous paper (44) which describes the detailed molecular composition of the HDL2-subfraction.

The size of simulated HDL particle decreased when moving from high to low HDL-C particle. This can be clearly seen from the radii of gyration (R_g) obtained from the simulation (Fig. S6). To estimate the hydrodynamic particle radii (R_H), we used the relation (R_g = sqrt(3/5) R_H) for a uniform sphere. The hydrodynamic diameter of high HDL-C particle was estimated to be 9.7 nm and for the low HDL-C it was 9.4 nm. As an indirect validation of our combined experimental and modeling strategy, the diameters calculated from simulations are in approximate agreement with our hydrodynamical particle diameters produced by electrophoresis, with the respective diameters being 9.9 and 9.0 nm when comparing high and low HDL-C particle profiles. One should note that the experimental size measurements include all HDL fractions and, thus, it is not directly comparable with our simulations. However, the results are within the right size range.

Interestingly, we registered by visual inspection that free cholesterol (FCho) molecules accumulated next to apoA-I proteins, indicating that free cholesterol molecules interact more favorably with apoA-I (Fig. 3B). This interaction may have important implications for the HDL metabolism as the cholesterol molecule concentration in the particles can modulate the conformational fluctuations of apoA-I or even change it. Indeed, studies with human apoA-I have shown that cholesterol decreases the adsorption of the apolipoprotein to a phospholipid monolayer, and the conformation of apoA-I varies from one lipid monolayer to another implying similar effects on the surface of HDL particles (45). It has been suggested that LCAT activity depends on the conformation of apoA-I (46, 47) and, thus, the cholesterol concentration could also be linked to the activation of LCAT and other HDL associated proteins. We therefore calculated the number of contacts between different lipid species and apoA-I (Fig. 3B). Furthermore, we produced root mean square fluctuation (RMSF) curves for apoA-I in different cases to characterize differences in

the local fluctuations (Fig. S7). The number of contacts increased between pairs SM-apoA-I, FCho-apoA-I and lysoPC-apoA-I when moving from low HDL-C to high HDL-C (Fig. 3B). In contrast, the number of contacts decreased between pairs ChoE-apoA-I, TG-apo-A-I and PC-apoA-I. However, only the cholesterol molecules accumulate preferably next to apoA-I proteins independent of cholesterol concentration. This suggests that the partitioning of cholesterol in HDL particles is entropy driven or that apoA-I possess high cholesterol affinity.

In order to study how the changes of lipid composition may affect the spatial distribution of specific lipid classes in HDL particles, we calculated the 3D radial distribution functions (RDFs) for different lipid components with respect to the center of mass of HDL particles (Figure 3c and Figs. S8-S10). Most notably, in low HDL-C the RDF of triacylglycerols indicated stronger prevalence of TG molecules at the surface (Fig. 3C) when the amount of cholesterol as well as other surface lipids decreases. The prevalence of cholesteryl esters did not change accordingly (Figs. S8-S10). Furthermore, we estimated the relative solubility TGs to the surface lipid monolayer by calculating the number of contacts between TGs and water phase, divided by the number of TG molecules (Fig. 3C). The relative solubility of TG increased with fewer surface lipids in the HDL particles.

DISCUSSION

Our study revealed marked differences in HDL lipidomic profiles as well as related clinical and biochemical characteristics between low and high HDL-C subjects. For the first time the molecular profiling of HDL (or other lipoprotein) particles was combined with dynamic structural modeling.

Such an approach allowed us to show that the lipid compositional changes also induce specific spatial distributions of lipids within the HDL particles.

The composition of lipid and apolipoprotein components is critical in maintaining normal HDL metabolism as well as its function. The HDL biosynthesis is very complex and is associated with major steric and lipid interaction changes in structural apolipoproteins such as apoA-I and apoA-II. ApoA-I which constitutes about 70 % of HDL protein content is present in almost all HDL particles (48). The lack of apoA-I gene is associated with low HDL-C levels in mice (49) and humans (50). This association is valid with our observations, with the low HDL-C subjects displaying lower levels of apoA-I as well as of apoA-II which constitutes approximately 20 % of HDL protein (51).

The combined information from the key metabolic regulators and lipid profiles may provide the basis for mechanistic links in HDL metabolism. In addition to apoA-I and apoA-II, the key metabolic regulators such as CETP, lecithin-cholesterol acyltransferase (LCAT), endothelial lipase (EL), hepatic lipase (HL) and PLTP play central role in continuous intravascular remodeling of HDL particles. ApoA-I acts as a co-factor for LCAT, the enzyme responsible for transforming the free cholesterol to cholesteryl ester in the core of HDL (52). As higher apoA-I levels closely associated with high HDL-C (48, 53) higher HDL-cholesteryl ester content was observed in high HDL-C subjects as compared to low HDL-C groups (Fig. 2D). The cholesteryl esters in HDL are continuously exchanged for TG in very low-density lipoprotein (VLDL) or chylomicrons in a process mediated by CETP activity. Our findings revealed increases of TG content in HDL particles in low HDL-C subjects, thus favoring the active ChoE/TG exchange process. TG

enrichment of HDL notably increases the ability of hepatic lipase to remodel these HDL-particles (54, 55) resulting in release of lipid-poor apoA-I and enhanced clearance of HDL *via* kidneys. It is known that cubilin acts as a receptor for the endocytosis of apoA-I in the kidney (56, 57).

Previous data suggest that in HDL particles the percentage of ChoEs of all surface lipids is between 13 – 27 % due to high tendency of ChoEs to locate on the surface and high concentration of ChoEs in these particles (58). Thus the known heteroexchange of ChoEs and TGs between HDL particles and apoB-100 containing lipoproteins is rational only if the acceptor and donor surfaces have different molar proportions of ChoEs and TG (58). In addition, higher concentration of TG molecules at the surface of HDL particles could give better chance for hepatic lipase to modify them.

The phospholipid content in HDL is regulated by another important factor PLTP (59). Mice with mild overexpression of PLTP did not reveal significant changes in HDL but those lacking PLTP have low HDL-C levels (60, 61). On the other, strong induction of PLTP expression strongly lowered total plasma HDL levels in a dose-dependent way due to accelerated fractional catabolism of HDL (62). In the present study PLTP activities did not differ between the two HDL groups suggesting that PLTP cannot be responsible for the low HDL levels. PLTP facilitates the transfer of phospholipids between lipoprotein particles and regulates both size and composition of HDL particles. The TG content of HDL has a major influence on PLTP-mediated size changes (63). The PLTP-mediated processes facilitate the transfer of surface remnants from lipolyzed triglyceride-rich lipoproteins to nascent HDL particles.

LCAT activation is influenced by the size of spherical HDL particles (64, 65). When the diameter of HDL particles increases the activity of LCAT enzyme is higher. Thus, we propose that the diameter of HDL particles is smaller in low HDL-C than in high HDL-C subjects. This lower activity could be somehow related to the conformation of apoA-I which could depend on the size and the lipid composition of the HDL particles. Interestingly, sphingomyelin content of HDL particles has been reported to affect LCAT activity (66, 67).

Intriguingly, as the conformation of apoA-I is different in each simulation performed in the present study, the notable differences in the fluctuations of apoA-I are possibly influenced by the conformation of the protein around the particle. Furthermore, as mentioned above the conformation of apoA-I could depend on the size and the lipid composition of the particles, especially the concentration of free cholesterol could be a critical determinant of apoA-I conformation as it was noticed by the simulations that cholesterol molecules favorably interact with apoA-Is. Presumably, the conformation and dynamical properties of apoA-I around the HDL particles are constantly changing since lipid and protein moieties vary. Therefore, producing any static or unified conformational data for apoA-Is around the native HDL particles is very challenging. However, by means of state-of-the-art coarse grained simulations we can now statistically define the most probable conformations of apoA-Is induced by the different molecular compositions of HDL particles. However, it is important to note that we should carry out more simulations to give more statistical power to our conformational changes seen in the present simulations. Our main aim in this study was to include apoA-I model that covers same surface area as native ApoA-I in the HDL particles and which is able to play a role in lipid partitioning. For this reason the protein tertiary structure differences seen in our simulation must be treated with certain

caution. Notably, the fatty acyl chain saturation/desaturation degree of phospholipids may affect the lateral movement of apoA-I on the particle surface.

Levels of polyunsaturated fatty acid containing ether phospholipids, including ethanolamine plasmalogens, were elevated in the HDL fractions of subjects with low HDL-C. Not much is known about the PC ether lipids in HDL, and in contrast to PE ether lipids (68), unambiguous structural characterization of these lipids is at present not possible. PE plasmalogens are known to be more abundant in HDL as compared to other lipoprotein fractions (7). Plasmalogens can serve as antioxidants against reactive oxygen species (ROS) and may prevent the oxidation of cholesterol (69). However, under oxidative stress the arachidonic acid containing plasmalogens, such as found in our study, become precursors of potent inflammatory mediators such as leukotrienes and hydroxyeicosatetraenoic acids (70). Together, the role of ether lipids in HDL function is likely complex and demands further investigation.

Sphingomyelin subspecies with different fatty acyl structures were higher in high HDL-C subjects. Interestingly, we also demonstrated that high HDL-C subjects have lower plasma TNFα as well as hs-CRP levels compared to that of low HDL-C subjects. This observation has important consequences when connected to sphingomyelin metabolism. A key event in atherogenesis is endothelial activation induced by a variety of stimuli such as tumor necrosis factor-alpha (TNF-α), resulting in the expression of various adhesion proteins (71). The expression of adhesion proteins on activated endothelial cells plays an essential role for the inflammatory processes in the pathogenesis of atherosclerosis (72). Elevated levels of TNF-α and hs-CRP inidicate a low-grade inflammation that could cause several downstream effects such as activation of sphingomyelinase

activities and affect the endothelial cell function and development of atherosclerosis. Connection to sphingomyelin metabolism is important and activation of acid sphingomyelinase has been extensively studied in the apst decade, and the enzyme is highly activated by TNF-α (73, 74) and might affect the lower SM content in HDL derived from low HDL subjects. This issue needs further studies such as *in vitro* cell cultures with the isolated HDL particles derived from low and high HDL cholesterol subjects in order to make further conclusions on particles physiological performance.

The marked differences in HDL lipid profiles reported here were obtained from the extreme ends of HDL-C levels in a large population cohort. Taken into account the relatively small sample size, the comparison of low *vs*. high HDL-C allowed us to minimize the effect of confounding factors and to study major differences. This approach thus helped us to extract the key compositional and structural features of HDL particles in the context of HDL-cholesterol which should be investigated further in other studies. To eliminate the confounding effect of apoA-I protein amount variation, the HDL particles were modeled based on the conservative assumption that the number of apoA-I molecules per particle is the same in high and low HDL-C group. One should thus keep in mind that our simulation model may not correctly represent the virtual structure of HDL in which different numbers of apoA-I are present. However, given that the observed increase of average apoA-I amount in high HDL-C subjects would only decrease the relative amount of triacylglycerols in high HDL-C subjects, the lipid spatial distribution differences between the high and low HDL-C are likely even bigger than estimated by our simulations.

In conclusion, we detected many important lipid compositional changes in subjects with high and low HDL-C. Elevated sphingomyelin in high HDL-C subjects confirms as well as provides additional evidence for the anti-inflammatory role of HDL. The lipid compositional changes also induced a surprising shift in the spatial distribution of triacylglycerols in subjects with low levels of HDL-C. The prevalence of TGs on the surface of HDL particles of these subjects may affect the hetero-exchange of core lipids by CETP as well as facilitate their modification by hepatic lipase. Finally, our study suggests that combining molecular profiling of lipoproteins with dynamic modeling of lipoprotein structure is a powerful new strategy that may help elucidate the complexity of systemic lipid metabolism as well as facilitate the efforts to invent novel therapeutic strategies.

Acknowledgements: The project was supported by the EU-funded project ETHERPATHS (FP7-KBBE-222639, http://www.etherpaths.org/), Finnish Heart Foundation, Sigrid Juselius Foundation, and Helsinki University Central Hospital Research Foundation. We are grateful to Pirkko Alha, Hannele Hilden, Ulla Lahtinen, Laura Lund, Jari Metso, Sari Nuutinen, Heli Nygren, Helinä Perttunen-Nio and Harri Rissanen for their excellent laboratory work and their help in data management.

REFERENCES

1. Gordon, T., W. P. Castelli, M. C. Hjortland, W. B. Kannel, and T. R. Dawber. 1977. High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. *Am. J. Med.* **62**: 707-714.

2. Chirovsky, D. R., V. Fedirko, Y. Cui, V. Sazonov, and P. Barter. 2009. Prospective studies on the relationship between high-density lipoprotein cholesterol and cardiovascular risk: a systematic review. *Eur. J. Cardiovasc. Prev. Rehabil* **16**: 404-423.

3. Cooney, M. T., A. Dudina, D. De Bacquer, L. Wilhelmsen, S. Sans, A. Menotti, G. De Backer, P. Jousilahti, U. Keil, T. Thomsen, P. Whincup, and I. M. Graham. 2009. HDL cholesterol protects against cardiovascular disease in both genders, at all ages and at all levels of risk. *Atherosclerosis* **206**: 611-616.

4. van Leuven, S. I., E. S. Stroes, and J. J. Kastelein. 2008. High-density lipoprotein: a fall from grace? *Ann. Med.* **40**: 584-593.

5. Watts, G. F., P. H. Barrett, and D. C. Chan. 2008. HDL metabolism in context: looking on the bright side. *Curr. Opin. Lipidol.* **19**: 395-404.

6. Vaisar, T., S. Pennathur, P. S. Green, S. A. Gharib, A. N. Hoofnagle, M. C. Cheung, J. Byun, S. Vuletic, S. Kassim, P. Singh, H. Chea, R. H. Knopp, J. Brunzell, R. Geary, A. Chait, X. Q. Zhao, K. Elkon, S. Marcovina, P. Ridker, J. F. Oram, and J. W. Heinecke. 2007. Shotgun proteomics implicates protease inhibition and complement activation in the antiinflammatory properties of HDL. *J. Clin. Invest.* **117**: 746-756.

7. Wiesner, P., K. Leidl, A. Boettcher, G. Schmitz, and G. Liebisch. 2009. Lipid profiling of FPLC-separated lipoprotein fractions by electrospray ionization tandem mass spectrometry. *J. Lipid Res.* **50**: 574-585.

8. Glass, C. K., and J. L. Witztum. 2001. Atherosclerosis: The Road Ahead. *Cell* **104**: 503-516.

9. Yancey, P. G., M. de la Llera-Moya, S. Swarnakar, P. Monzo, S. M. Klein, M. A. Connelly, W. J. Johnson, D. L. Williams, and G. H. Rothblat. 2000. High Density Lipoprotein Phospholipid Composition Is a Major Determinant of the Bi-directional Flux and Net Movement of Cellular Free Cholesterol Mediated by Scavenger Receptor BI. *J. Biol. Chem.* **275**: 36596-36604.

10. Brewer, H. B., Jr. 2004. Increasing HDL Cholesterol Levels. *N. Engl. J. Med.* **350**: 1491-1494.

11. Nakanishi, S., R. Vikstedt, S. Soderlund, M. Lee-Rueckert, A. Hiukka, C. Ehnholm, M. Muilu, J. Metso, J. Naukkarinen, L. Palotie, P. T. Kovanen, M. Jauhiainen, and M.-R. Taskinen. 2009. Serum, but not monocyte macrophage foam cells derived from low HDL-C subjects, displays reduced cholesterol efflux capacity. *J. Lipid Res.* **50**: 183-192.

12. Wang, N., D. Lan, W. Chen, F. Matsuura, and A. R. Tall. 2004. ATP-binding cassette transporters G1 and G4 mediate cellular cholesterol efflux to high-density lipoproteins. *Proc. Natl. Acad. Sci. USA.* **101**: 9774-9779.

13. Fournier, N., J.-L. Paul, V. Atger, A. Cogny, T. Soni, M. de la Llera-Moya, G. Rothblat, and N. Moatti. 1997. HDL Phospholipid Content and Composition as a Major Factor Determining Cholesterol Efflux Capacity From Fu5AH Cells to Human Serum. *Arterioscler. Thromb. Vasc. Biol.* **17**: 2685-2691.

14. Davidson, W. S., K. L. Gillotte, S. Lund-Katz, W. J. Johnson, G. H. Rothblat, and M. C. Phillips. 1995. The Effect of High Density Lipoprotein Phospholipid Acyl Chain Composition on the Efflux of Cellular Free Cholesterol. *J. Biol. Chem.* **270**: 5882-5890.

15. Vance, D. E., and J. E. Vance. 2008. Biochemistry of lipids,lipoproteins and membranes Elsevier, Hungary.

16. Kotronen, A., V. Velagapudi, L. Yetukuri, J. Westerbacka, R. Bergholm, K. Ekroos, J. Makkonen, M. R. Taskinen, M. Oresic, and H. Yki-Järvinen. 2009. Serum saturated fatty acids containing triacylglycerols are better markers of insulin resistance than total serum triacylglycerol concentrations. *Diabetologia* **52**: 684-690.

17. Han, X., and R. W. Gross. 2005. Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. *Mass Spectrom. Rev.* **24**: 367-412.

18. Oresic, M., V. A. Hänninen, and A. Vidal-Puig. 2008. Lipidomics: a new window to biomedical frontiers. *Trends Biotechnol.* **26**: 647-652.

19. Aromaa, A., and S. Koskinen. 2004. Health and Functional Capacity in Finland: Baseline Results of the Health 2000 Health Examination Survey. National Public Health Institute, Helsinki, Finland.

20. Taskinen, M. R., T. Kuusi, E. Helve, E. A. Nikkila, and H. Yki-Jarvinen. 1988. Insulin therapy induces antiatherogenic changes of serum lipoproteins in noninsulin-dependent diabetes. *Arteriosclerosis* **8**: 168-177.

21. Blanche, P. J., E. L. Gong, T. M. Forte, and A. V. Nichols. 1981. Characterization of human high-density lipoproteins by gradient gel electrophoresis. *Biochim Biophys Acta* **665**: 408-419.

22. Perusse, M., A. Pascot, J. P. Despres, C. Couillard, and B. Lamarche. 2001. A new method for HDL particle sizing by polyacrylamide gradient gel electrophoresis using whole plasma. *J. Lipid Res.* **42**: 1331-1334.

23. Vakkilainen, J., M. Jauhiainen, K. Ylitalo, I. O. Nuotio, J. S. Viikari, C. Ehnholm, and M. R. Taskinen. 2002. LDL particle size in familial combined hyperlipidemia: effects of serum lipids, lipoprotein-modifying enzymes, and lipid transfer proteins. *J. Lipid Res.* **43**: 598-603.

24. Friedewald, W. T., R. I. Levy, and D. S. Fredrickson. 1972. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* **18**: 499-502.

25. Siggins, S., M. Jauhiainen, V. M. Olkkonen, J. Tenhunen, and C. Ehnholm. 2003. PLTP secreted by HepG2 cells resembles the high-activity PLTP form in human plasma. *J. Lipid Res.* **44**: 1698-1704.

26. Damen, J., J. Regts, and G. Scherphof. 1982. Transfer of [14C]phosphatidylcholine between liposomes and human plasma high density lipoprotein. Partial purification of a transfer-stimulating plasma factor using a rapid transfer assay. *Biochim Biophys Acta* **712**: 444-452.

27. Jauhiainen, M., and C. Ehnholm. 2005. Determination of human plasma phospholipid transfer protein mass and activity. *Methods* **36**: 97-101.

28. Siggins, S., M. Karkkainen, J. Tenhunen, J. Metso, E. Tahvanainen, V. M. Olkkonen, M. Jauhiainen, and C. Ehnholm. 2004. Quantitation of the active and low-active forms of human plasma phospholipid transfer protein by ELISA. *J. Lipid Res.* **45**: 387-395.

29. Groener, J. E., R. W. Pelton, and G. M. Kostner. 1986. Improved estimation of cholesteryl ester transfer/exchange activity in serum or plasma. *Clin. Chem.* **32**: 283-286.

30. Kleemola, P., R. Freese, M. Jauhiainen, R. Pahlman, G. Alfthan, and M. Mutanen. 2002. Dietary determinants of serum paraoxonase activity in healthy humans. *Atherosclerosis* **160**: 425-432.

31. Geladi, P., and B. R. Kowalski. 1986. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **185**: 1-17.

32. Matthew Barker, W. R. 2003. Partial least squares for discrimination. *J. Chemometr.* **17**: 166-173.

33. Wise, B. M., N. B. Gallagher, R. Bro, J. M. Shaver, W. Windig, and J. S. Koch. 2005. PLS Toolbox 3.5 for use with Matlab. Manson, WA, Eigenvector Research Inc.

34. Wold, S., K. Esbensen, and P. Geladi. 1987. Principal Component analysis. *Chemometr. Intell. Lab. Syst.* **2**: 37-52.

35. Borhani, D., D. Rogers, J. Engler, and C. Brouillette. 1997. Crystal structure of truncated human apolipoprotein A-I suggests a lipid-bound conformation. *Proc. Natl. Acad. Sci. USA* **94**: 12291-12296.

36. Segrest, J., M. Jones, A. Klon, C. Sheldahl, M. Hellinger, H. De Loof, and S. Harvey. 1999. A Detailed Molecular Belt Model for Apolipoprotein A-I in Discoidal High Density Lipoprotein. *J. Biol. Chem.* **274**: 31755-31758.

37. Okon, M., P. Frank, Y. Marcel, and R. Cushley. 2002. Heteronuclear NMR studies of human serum apolipoprotein A-I: Part I. Secondary structure in lipid-mimetic solution. *FEBS Lett.* **517**: 139-143.

38. Catte, A., J. Patterson, D. Bashtovyy, M. Jones, L. Gu F, L., A. Rampioni, D. Sengupta, T. Vuorela, P. Niemelä, M. Karttunen, S. Marrink, I. Vattulainen, and J. Segrest. 2008. Structure of Spheroidal HDL Particles Revealed by Combined Atomistic and Coarse-Grained Simulations. *Biophys. J.* **94**: 2306-2319.

39. Monticelli, L., S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink. 2008. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J Chem Theory Comput* **4**: 819-834.

40. Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl. 2008. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4**: 435-447.

41. Marrink, S., J. Risselada, S. Yefimov, P. Tieleman, and A. de Vries. 2007. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **111**: 7812-7824.

42. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**: 3684-3690.

43. Pietiläinen, K. H., M. Sysi-Aho, A. Rissanen, T. Seppänen-Laakso, H. Yki-Järvinen, J. Kaprio, and M. Orešic. 2007. Acquired Obesity Is Associated with Changes in the Serum Lipidomic Profile Independent of Genetic Effects- A Monozygotic Twin Study. *PLoS ONE* **2**: e218.

44. Lund-Katz, S., L. Liu, S. T. Thuahnai, and M. C. Phillips. 2003. High density lipoprotein structure. *Front. Biosci.* **8**: d1044-1054.

45. Ibdah, J. A., and M. C. Phillips. 2002. Effects of lipid composition and packing on the adsorption of apolipoprotein A-I to lipid monolayers. *Biochemistry* **27**: 7155-7162.

46. Sorci-Thomas, M. G., L. Curtiss, J. S. Parks, M. J. Thomas, M. W. Kearns, and M. Landrum. 1998. The Hydrophobic Face Orientation of Apolipoprotein A-I Amphipathic Helix Domain 143-164 Regulates Lecithin:Cholesterol Acyltransferase Activation. *J. Biol. Chem.* **273**: 11776-11782.

47. Cho, K.-H., D. M. Durbin, and A. Jonas. 2001. Role of individual amino acids of apolipoprotein A-I in the activation of lecithin:cholesterol acyltransferase and in HDL rearrangements. *J. Lipid Res.* **42**: 379-389.

48. Rader, D. J. 2006. Molecular regulation of HDL metabolism and function: implications for novel therapies. *J. Clin. Invest.* **116**: 3090-3100.

49. Williamson, R., D. Lee, J. Hagaman, and N. Maeda. 1992. Marked reduction of high density lipoprotein cholesterol in mice genetically modified to lack apolipoprotein A-I. *Proc. Natl. Acad. Sci. USA* **89**: 7134-7138.

50. Schaefer, E., W. Heaton, M. Wetzel, and H. Brewer, Jr. 1982. Plasma apolipoprotein A-1 absence associated with a marked reduction of high density lipoproteins and premature coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.* **2**: 16-26.

51. Weng, W., and J. L. Breslow. 1996. Dramatically decreased high density lipoprotein cholesterol, increased remnant clearance, and insulin hypersensitivity in apolipoprotein A-II knockout mice suggest a complex role for apolipoprotein A-II in atherosclerosis susceptibility. *Proc. Natl. Acad. Sci. USA* **93**: 14788-14794.

52. Nakamura, Y., L. Kotite, Y. Gan, T. A. Spencer, C. J. Fielding, and P. E. Fielding. 2004. Molecular Mechanism of Reverse Cholesterol Transport: Reaction of Pre-beta-Migrating High-Density Lipoprotein with Plasma Lecithin/Cholesterol Acyltransferase. *Biochemistry* **43**: 14811-14820.

53. Watanabe, H., S. Soderlund, A. Soro-Paavonen, A. Hiukka, E. Leinonen, C. Alagona, R. Salonen, T.-P. Tuomainen, C. Ehnholm, M. Jauhiainen, and M.-R. Taskinen. 2006. Decreased High-Density Lipoprotein (HDL) Particle Size, Pre{beta}-, and Large HDL Subspecies Concentration in Finnish Low-HDL Families: Relationship With Intima-Media Thickness. *Arterioscler Thromb Vasc Biol* **26**: 897-902.

54. Barter, P. J. 2002. Hugh Sinclair Lecture: The regulation and remodelling of HDL by plasma factors. *Atherosclerosis Supp.* **3**: 39-47.

55. Lamarche, B. Â. t., K. D. Uffelman, A. Â. Ã. Â. Carpentier, J. S. Cohn, G. Steiner, P. H. Barrett, and G. F. Lewis. 1999. Triglyceride enrichment of HDL enhances in vivo metabolic clearance of HDL apo A-I in healthy men. *J. Clin. Investi.* **103**: 1191-1199.

56. Hammad, S. M., S. Stefansson, W. O. Twal, C. J. Drake, P. Fleming, A. Remaley, H. B. Brewer, and W. S. Argraves. 1999. Cubilin, the endocytic receptor for intrinsic factor-vitamin B12 complex, mediates high-density lipoprotein holoparticle endocytosis. *Proc. Natl. Acad. Sci. USA* **96**: 10158-10163.

57. Kozyraki, R., J. Fyfe, M. Kristiansen, C. Gerdes, C. Jacobsen, S. Cui, E. I. Christensen, M. Aminoff, A. de la Chapelle, R. Krahe, P. J. Verroust, and S. K. Moestrup. 1999. The intrinsic factor-vitamin B12 receptor, cubilin, is a high-affinity apolipoprotein A-I receptor facilitating endocytosis of high-density lipoprotein. *Nat. Med.* **5**: 656-661.

58. Kumpula, L. S., J. M. Kumpula, M.-R. Taskinen, M. Jauhiainen, K. Kaski, and M. Ala-Korpela. 2008. Reconsideration of hydrophobic lipid distributions in lipoprotein particles. *Chem. Phys. Lipids* **155**: 57-62.

59. Huuskonen, J., V. M. Olkkonen, M. Jauhiainen, and C. Ehnholm. 2001. The impact of phospholipid transfer protein (PLTP) on HDL metabolism. *Atherosclerosis* **155**: 269-281.

60. Jiang, X., O. L. Francone, C. Bruce, R. Milne, J. Mar, A. Walsh, J. L. Breslow, and A. R. Tall. 1996. Increased prebeta-high density lipoprotein, apolipoprotein AI, and phospholipid in mice expressing the human phospholipid transfer protein and human apolipoprotein AI transgenes. *J. Clin. Invest.* **98**: 2373-2380.

61. Jiang, X.-c., C. Bruce, J. Mar, M. Lin, Y. Ji, O. L. Francone, and A. R. Tall. 1999. Targeted mutation of plasma phospholipid transfer protein gene markedly reduces high-density lipoprotein levels. *J. Clin. Invest.* **103**: 907-914.

62. Föger, B., S. Santamarina-Fojo, R. Shamburek, C. Parrot, G. Talley, and H. Brewer. 1997. Plasma Phospholipid Transfer Protein. Adenovirus-mediated overexpression in mice leads to decreased plasma High Density Lipoprotein (HDL) and enhanced hepatic uptake of phospholipids and cholesteryl esters from HDL. *J. Biol. Chem.* **272**: 27393-27400.

63. Rye, K.-A., M. Jauhiainen, P. J. Barter, and C. Ehnholm. 1998. Triglyceride-enrichment of high density lipoproteins enhances their remodelling by phospholipid transfer protein. *J. Lipid Res.* **39**: 613-622.

64. Cavigiolio, G., B. Shao, E. G. Geier, G. Ren, J. W. Heinecke, and M. N. Oda. 2008. The Interplay between Size, Morphology, Stability, and Functionality of High-Density Lipoprotein. SubclassesÃ‚Â *Biochemistry* **47**: 4770-4779.

65. Curtiss, L. K., D. J. Bonnet, and K.-A. Rye. 2000. The Conformation of Apolipoprotein A-I in High-Density Lipoproteins Is Influenced by Core Lipid Composition and Particle Size: A Surface Plasmon Resonance Study. *Biochemistry* **39**: 5712-5721.

66. Bolin, D. J., and A. Jonas. 1996. Sphingomyelin Inhibits the Lecithin-Cholesterol Acyltransferase Reaction with Reconstituted High Density Lipoproteins by Decreasing Enzyme Binding. *J. Biol. Chem.* **271**: 19152-19158.

67. Subbaiah, P. V., P. Horvath, and S. B. Achar. 2006. Regulation of the Activity and Fatty Acid Specificity of Lecithin-Cholesterol Acyltransferase by Sphingomyelin and Its Metabolites, Ceramide and Ceramide Phosphate. *Biochemistry* **45**: 5029-5038.

68. Zemski Berry, K. A., and R. C. Murphy. 2004. Electrospray ionization tandem mass spectrometry of glycerophosphoethanolamine plasmalogen phospholipids. *J. Am. Soc. Mass Spectrom.* **15**: 1499-1508.

69. Maeba, R., and N. Ueta. 2003. Ethanolamine plasmalogens prevent the oxidation of cholesterol by reducing the oxidizability of cholesterol in phospholipid bilayers. *J. Lipid Res.* **44**: 164-171.

70. Murphy, R. C. 2001. Free-radical-induced oxidation of arachidonoyl plasmalogen phospholipids: antioxidant mechanism and precursor pathway for bioactive eicosanoids. *Chem. Res. Toxicol.* **14**: 463-472.

71. Xia, P., J. R. Gamble, K.-A. Rye, L. Wang, C. S. T. Hii, P. Cockerill, Y. Khew-Goodall, A. G. Bert, P. J. Barter, and M. A. Vadas. 1998. Tumor necrosis factor-alpha induces adhesion molecule expression through the sphingosine kinase pathway. *Proc. Natl. Acad. Sci. USA.* **95**: 14196-14201.

72. Ross, R. 1999. Atherosclerosis -- An Inflammatory Disease. *N. Engl. J. Med.* **340**: 115-126.

73. García-Ruiz, C., A. Colell, M. Marí, A. Morales, M. Calvo, C. Enrich, and J. C. Fernández-Checa. 2003. Defective TNF-alpha-mediated hepatocellular apoptosis and liver damage in acidic sphingomyelinase knockout mice. *J Clin Invest* **111**: 197-208.

74. Schütze, S., K. Potthoff, T. Machleidt, D. Berkovic, K. Wiegmann, and M. Krönke. 1992. TNF activates NF-[kappa]B by phosphatidylcholine-specific phospholipase C-induced "Acidic" sphingomyelin breakdown. *Cell* **71**: 765-776.

Figure legends

Fig. 1. (A) Partial least squares discriminant analysis (PLS/DA) of lipidomic profiles for low HDL-C and high HDL-C subjects. PLS/DA scores plot with two different HDL-C groups are marked. Two latent variables were used (Q^2=51%). (B) Hierarchical clustering on most important variable importance in the projection (VIP) variables and samples in the heat map reflecting fold changes of lipids relative to mean intensity within the low HDL-C group. Bars show fold changes reflecting mean intensity of top VIP variables in the high HDL-C group relative to mean intensity within the low HDL-C group.

Fig. 2. Box plots of the most abundant lipids within the TG, lysoPC, SM, ChoE and ethanolamine plasmalogen (PEp) classes. Concentrations are shown in μmol/l [lipid] / mg/dl [apoA-I]. High, High HDL-C; Low, Low HDL-C.

Fig. 3. Coarse grained simulations of HDL particles reconstituted based on lipidomics data. (A) Snapshots from the end of high and low HDL-C simulations (8 μs). Apo-AIs are colored with red and green, cholesterol molecules are yellow and all other lipids grey. Water phase was removed from the snapshots for clarity. Middle snapshot demonstrates how the cholesterol molecules are localized next to and under apoA-Is in high HDL-C simulation. (B) The number of contacts between apoA-Is and different lipid beads in each simulation (error bars indicate ± SD). The number of contacts was not normalized with number of different lipid constituents. (C) Radial distribution function for TG molecules g(r) with respect to the center of mass (COM) of HDL particle. When surface lipid concentration decreases more TGs are able to penetrate the surface

monolayer which is seen as a formation of higher plateaus near 3.5 nm. The number of contacts between TG and water beads (per TG) in different simulations is listed in the inset.

.

Table 1. Clinical and biochemical characteristics of the study subjects. IQR; interquartile range. **P*-value from Mann-Whitney *U* test. **apoA-II measurements of three subjects both in low and high HDL-C groups are not available and hence calculations are based on measurements from the remaining subjects

	Low-HDL subjects (median (IQR))	**High-HDL subjects (median (IQR))**	***P*-value***
N (men/women)**	24(12/12)	23(12/11)	
Age (years)	53 (51-56)	54 (50-60)	0.564
Body mass index (kg/m^2)	27.9 (24.3-31.5)	22.8 (21.4-24.6)	<0.001
Systolic blood pressure (mmHg)	131 (118-146)	132 (121-147)	0.647
Diastolic blood pressure (mmHg)	81 (75-89)	81 (75-87)	0.882
Total cholesterol (mmol/l)	5.15 (4.55-5.60)	5.80 (5.30-6.10)	<0.001
HDL cholesterol (mmol/l)	0.94 (0.86-1.10)	2.52 (2.12-2.61)	<0.001
LDL cholesterol (mmol/l)	3.30 (2.87-3.67)	3.01 (2.68-3.55)	0.225
Insulin (mU/l)	9.15 (7.05-10.93)	6.00 (5.00-7.90)	0.0015
Triglycerides (mmol/l)	1.80 (1.23-2.25)	0.70 (0.60-0.90)	<0.001
apoA-I (mg/dl)	134 (121-144)	222 (206-234)	<0.001
apoA-II (mg/dl)	28.0 (27.0-34.5)	39.0 (35.0-45.0)	<0.001
apoB (mg/dl)	124.5 (106.3-135.0)	99.0 (84.0-109.0)	<0.001
PLTP activity (nmol/ml/h)	4765 (4109-5549)	5304 (4798-5810)	0.058
PLTP mass (µg/ml)	7.5 (6.4-9.2)	8.6 (6.8-9.5)	0.328
CETP activity (nmol/ml/h)	28 (25-33)	31 (27-38)	0.250
TNF-α (ng/l)	6.5 (4.5-8.2)	4.7 (4.1-5.8)	0.038
IL-6 (Interleukin 6)(ng/l)	1.6 (1.0-2.1)	1.1 (0.8-2.0)	0.163

hs-CRP (mg/l)	1.4 (0.8-2.7)	0.7 (0.4-1.2)	0.012
HDL size (nm)	9.0 (8.8-9.2)	9.9 (9.7-10.2)	<0.001

.

Table 2.

(a) The lipid compositions (number of molecules per HDL particle) used in simulations.

	apoA-I	SM	PC	FCho	ChoE	TG	lysoPC
normal-HDL-C	2	18	109	50	90	19	10
low-HDL-C	2	13	109	25	81	24	5
high-HDL-C	2	23	109	75	99	14	15

(b) The average concentrations of the lipid classes as determined by lipidomics, which used to determine the lipid compositions in simulations shown in (a). Units: μmol/l [lipid] / mg/dl [apoA-I].

	SM	HDL-C	ChoE	TG	lysoPC
low-HDL-C	0.057 ± 0.003	7.545 ± 0.200	0.105 ± 0.009	0.436 ± 0.024	0.014 ± 0.007
high HDL-C	0.082 ± 0.004	10.928 ± 0.254	0.132 ± 0.020	0.244 ± 0.036	0.022 ± 0.002

Figure 1

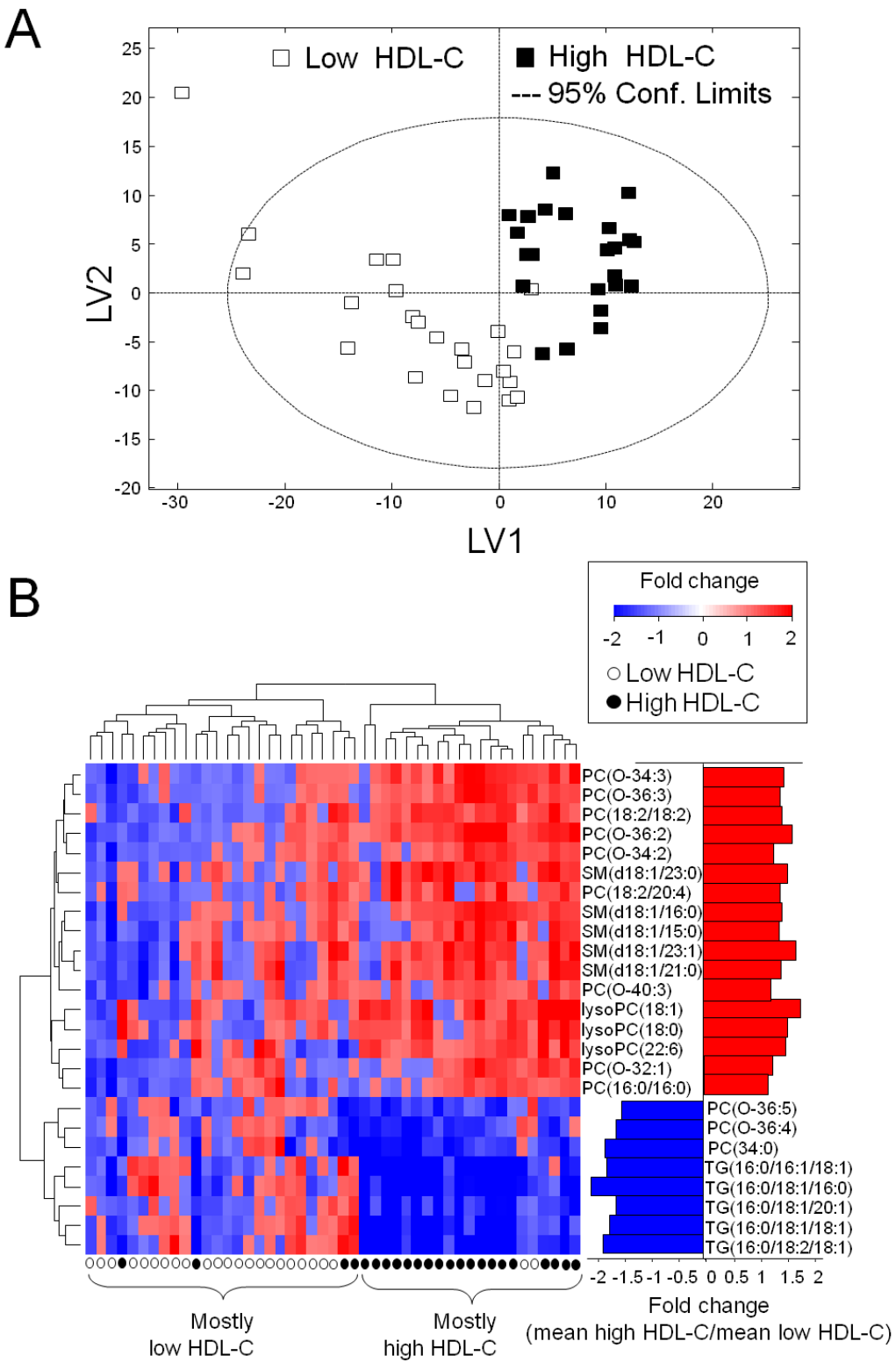


Figure 2

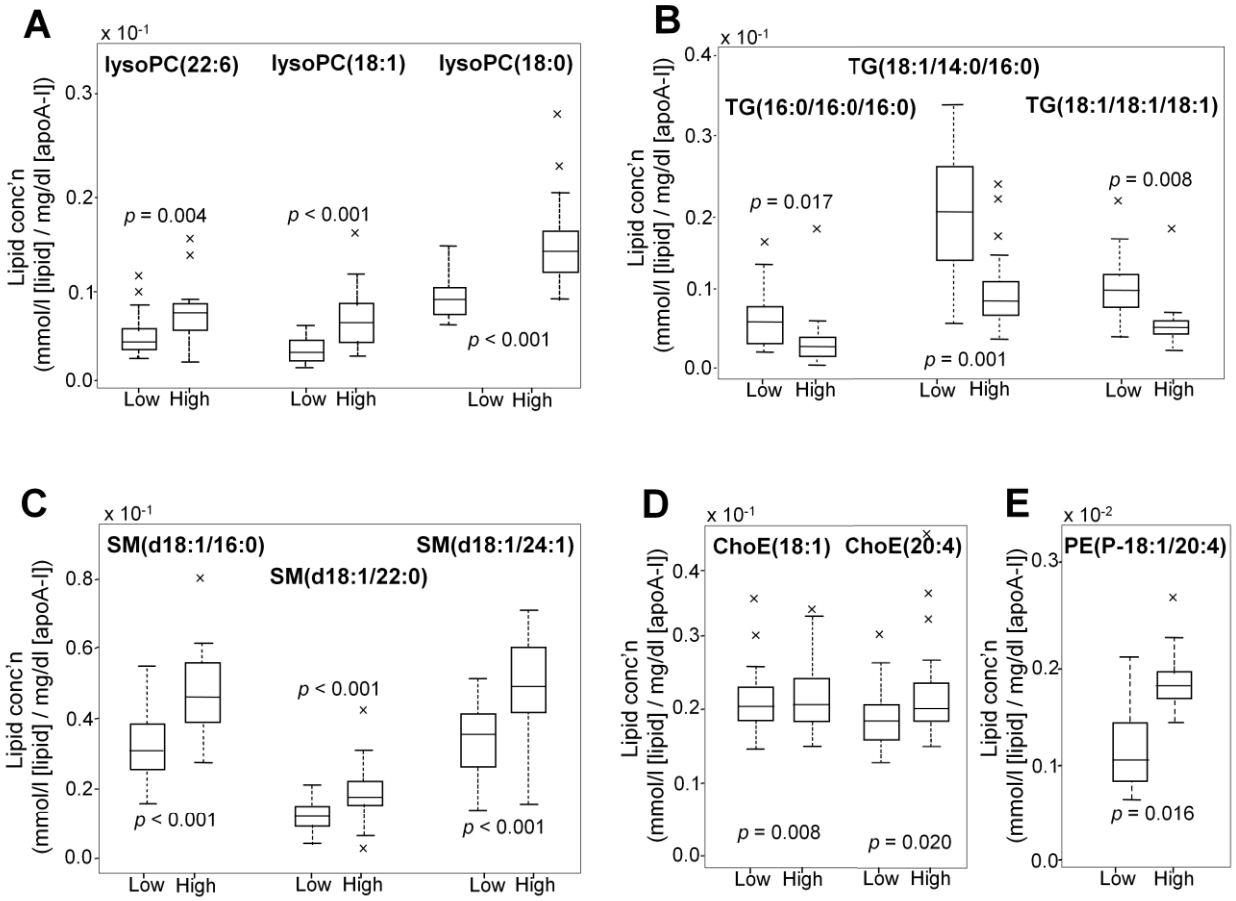
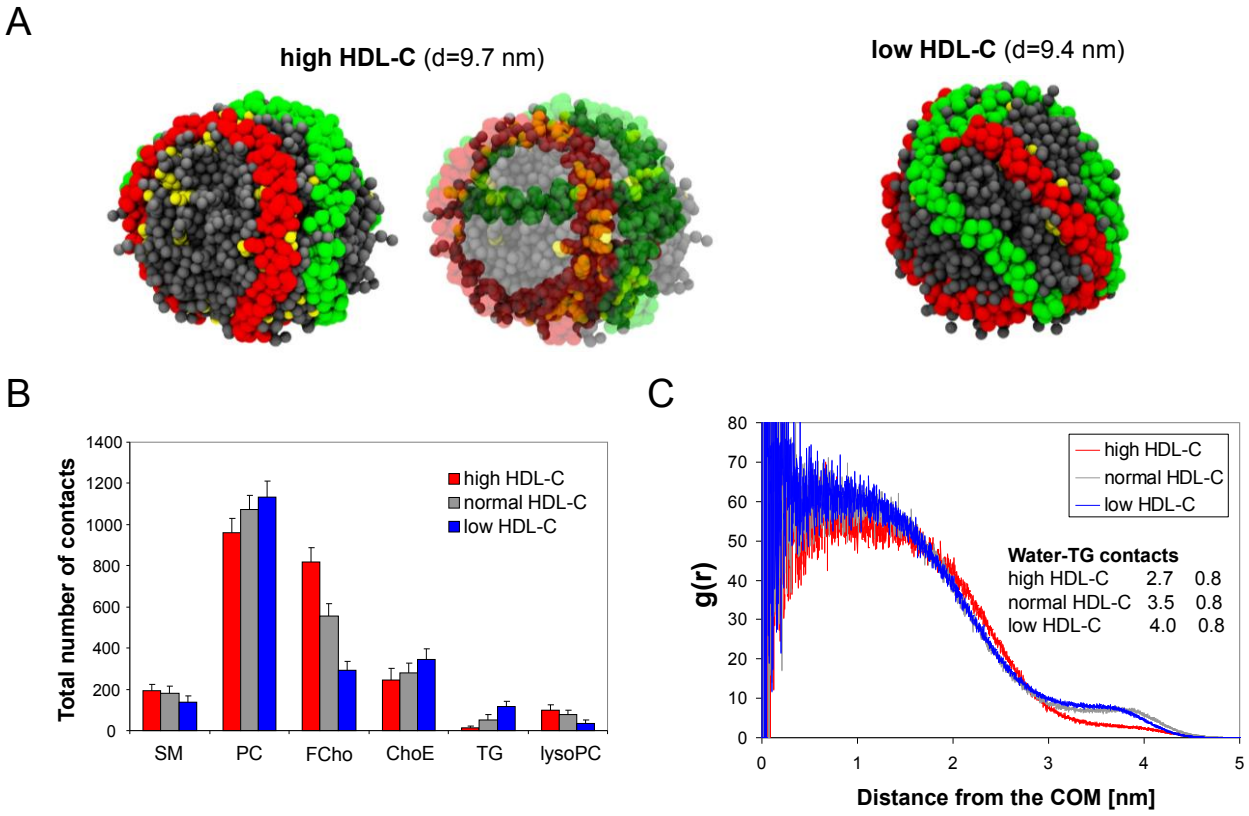


Figure 3



PUBLICATION IV

Adaptation and failure of pancreatic beta cells in murine models with different degrees of metabolic syndrome

In: Desease Models & Mechanisms 2009.
Vol. 2, pp. 582–592.

RESEARCH ARTICLE

Disease Models & Mechanisms 2, 582-592 (2009) doi:10.1242/dmm.003251
Published by The Company of Biologists 2009

Adaptation and failure of pancreatic β cells in murine models with different degrees of metabolic syndrome

Gema Medina-Gomez[1,4,*], Laxman Yetukuri[2], Vidya Velagapudi[2], Mark Campbell[1], Margaret Blount[1], Mercedes Jimenez-Linan[3], Manuel Ros[1], Matej Orešič[2] and Antonio Vidal-Puig[1,*]

SUMMARY

The events that contribute to the expansion of β-cell mass and enhanced β-cell function in insulin-resistant states have not been elucidated fully. Recently, we showed that β-cell adaptation failed dramatically in adult, insulin-resistant POKO mice, which contrasts with the appropriate expansion of β cells in their *ob/ob* littermates. Thus, we hypothesised that characterisation of the islets in these mouse models at an early age should provide a unique opportunity to: (1) identify mechanisms involved in sensing insulin resistance at the level of the β cells, (2) identify molecular effectors that contribute to increasing β-cell mass and function, and (3) distinguish primary events from secondary events that are more likely to be present at more advanced stages of diabetes. Our results define the POKO mouse as a model of early lipotoxicity. At 4 weeks of age, it manifests with inappropriate β-cell function and defects in proliferation markers. Other well-recognised pathogenic effectors that were observed previously in 16-week-old mice, such as increased reactive oxygen species (ROS), macrophage infiltration and endoplasmic reticulum (ER) stress, are also present in both young POKO and young *ob/ob* mice, indicating the lack of predictive power with regards to the severity of β-cell failure. Of interest, the relatively preserved lipidomic profile in islets from young POKO mice contrasted with the large changes in lipid composition and the differences in the chain length of triacylglycerols in the serum, liver, muscle and adipose tissue in adult POKO mice. Later lipotoxic insults in adult β cells contribute to the failure of the POKO β cell. Our results indicate that the rapid development of insulin resistance and β-cell failure in POKO mice makes this model a useful tool to study early molecular events leading to insulin resistance and β-cell failure. Furthermore, comparisons with *ob/ob* mice might reveal important adaptive mechanisms in β cells with either therapeutic or diagnostic potential.

INTRODUCTION

The common forms of insulin-resistant type 2 diabetes typically involve an early phase characterised by insulin resistance, increased insulin secretion and a progressive expansion of β-cell mass. In response to insulin resistance, the islet machinery is challenged to secrete sufficient insulin to maintain euglycaemia. This requires robust adaptation that causes metabolic stress or an allostatic load on the physiological mechanisms that control insulin secretion and β-cell mass proliferation (Stumvoll et al., 2003). At some point, probably as a consequence of genetic vulnerability to metabolic stress, the balance between impaired insulin signalling and increased secretion of insulin fails, leading to hyperglycaemia and dyslipidaemia. These toxic metabolic disturbances increase the allostatic load on the system, ultimately causing a major collapse of the adaptive mechanisms and β-cell function (Stumvoll et al., 2003).

Studies performed in states of frank diabetes that aim to identify mechanisms involved in β-cell failure typically fail to distinguish early pathogenic mechanisms from either adaptive homeostatic responses or secondary pathogenic effects. In some cases, the same adaptive mechanism might exert opposite metabolic effects depending on how far advanced the β-cell failure is. This indicates that the mechanisms that provide an initial, positive, adaptive physiological response might become toxic at later stages when acting at high intensity and/or for extended periods of time. This concept is known as hormesis. For example, although the short-term exposure of islets to fatty acids promotes insulin secretion (Prentki et al., 2002; Roduit et al., 2004), chronic exposure to, and/or high concentrations of, fatty acids leads to β-cell failure (Prentki and Corkey, 1996; Bollheimer et al., 1998; Prentki et al., 2002). Similarly, there is evidence that reactive oxygen species (ROS), which are traditionally considered to be agents of molecular stress and damage (Brownlee, 2003; Lowell and Shulman, 2005), might, at small doses and for relatively short periods of time, provide a physiological signal that couples oxidation of glucose to insulin secretion (Li et al., 2006). This specific example is known as mitohormesis or mitochondrial hormesis (Schulz et al., 2007). Our hypothesis is that, in response to insulin resistance, specific responses in pancreatic β cells acutely improve their capacity to secrete insulin and trigger the expansion of β-cell mass. Although in the short term these adaptive mechanisms are efficient at maintaining carbohydrate homeostasis, when they are activated chronically, the same mechanisms might result in an excessive allostatic load that, ultimately, leads to collapse of the metabolic regulatory networks that control insulin secretion. Accumulating evidence indicates that diagnostic and therapeutic interventions at this late stage, when the homeostasis of the β-cell functional network is already lost, are inefficient, hence our focus on the early stages of the evolutive process where the dynamics of the

[1]University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK
[2]VTT Technical Research Centre of Finland, Tietotie 2, Espoo, P.O. Box1500, FIN-02044 VTT, Finland
[3]Department of Histopathology, Addenbrooke's Hospital, University of Cambridge, Cambridge, CB2 0QQ
[4]Present address: Departamento de Bioquímica, Fisiología y Genética Molecular, Universidad Rey Juan Carlos, Facultad de Ciencias de la Salud, Avda. de Atenas s/n, 28922 Alcorcón, Madrid, Spain
*Authors for correspondence (e-mail: gema.medina@urjc.es; ajv22@cam.ac.uk)

functional network can still be recognised and potentially reversed.

We recently generated a murine model, the POKO mouse, which was obtained by crossing a peroxisome proliferator-activated receptor gamma 2 (PPARγ2) knockout (KO) mouse into a genetically obese, insulin-resistant *ob/ob* background (Medina-Gomez et al., 2007). POKO and *ob/ob* mice have a similar positive energy balance (hyperphagia and energy dissipation), but because POKO mice lack PPARγ2 they are unable to expand adipose tissue and, as a result, they become lipotoxic and markedly more insulin resistant than their *ob/ob* littermates. By the age of 16 weeks, POKO mice show severe β-cell failure and hyperglycaemia. Another difference between the two mouse models is the lack of expansion of β cells in POKO mice, which contrasts with the markedly increased expansion of β-cell mass in insulin-resistant *ob/ob* littermates. Of interest, 4-week-old *ob/ob* and POKO mice have a similar islet number and morphology, which allowed us to investigate the mechanisms that might predict subsequent differences in metabolic profiles and expansion of β-cell mass in the two models.

The aim of our study was to investigate the early molecular regulatory networks that might account for the later differences in β-cell mass and function in insulin-resistant *ob/ob* and POKO mice. We speculated that these studies should validate the young POKO mouse as a model that could be used to study the events leading to β-cell failure and that would be suitable for future systems biology studies. This experimental paradigm also offers the possibility of identifying novel key events and molecular effectors that are involved in the adaptation of β cells to insulin resistance. Here, we show that isolated islets from mice as young as 4-5 weeks old reveal signs of inappropriate adaptive β-cell function and lipid signalling. Also, we provide evidence that other pathogenic effectors that are observed characteristically in advanced stages of disease, such as increased ROS production, inflammation, macrophage infiltration and endoplasmic reticulum (ER) stress, are already present at 4 weeks of age, which indicates their involvement in early pathogenic changes. However, our data also indicate that whereas these pathogenic mechanisms are present in β cells from early stages, they do not predict the rapid development of hyperglycaemia in POKO mice. Our data also indicate that the altered lipid composition of islets is a relatively late event in the context of lipotoxicity, which is preceded by lipid changes in the serum, liver, adipose tissue and muscle.

RESULTS

Differences in β-cell mass and function in *ob/ob* and POKO mice

We have reported previously that, in contrast to *ob/ob* mice, POKO mice do not expand their β-cell mass in response to insulin resistance (Medina-Gomez et al., 2007). As a result, adult POKO mice suffer metabolic collapse associated with severe β-cell failure that evolves over a relatively short period of 16 weeks. This contrasts with the robust expansion of β cells and the milder metabolic disturbances that are observed in *ob/ob* littermate controls, despite their marked obesity and insulin resistance. Interestingly, at the age of 4 weeks, the morphology of islets from POKO and *ob/ob* mice was indistinguishable, with the staining for insulin and glucagon being similar in both types of mice (Medina-Gomez et al., 2007).

Our initial studies in isolated islets from 4-week-old mice revealed that, at a low glucose concentration (2.5 mM), basal insulin secretion is maintained in POKO and *ob/ob* mice when compared with wild-type (WT) controls (Fig. 1A). Insulin secretion after 25 mM of glucose was stimulated significantly, but to a similar extent, in the insulin-resistant *ob/ob* and POKO mice. This indicates that, at this early age, the differences in insulin secretory defects between POKO and *ob/ob* mice are not established fully.

To validate the POKO model, we initially investigated the expression of genes involved in β-cell proliferation in young (5-week-old) C57BL/6 female WT, PPARγ2 KO, *ob/ob* and POKO mice. Insulin receptor substrate 2 (IRS2) (Fig. 1B) has been shown previously to be a key mediator of β-cell proliferation and to exert a protective role against β-cell glucotoxicity (Choudhury et al., 2005; Wang et al., 2005; Cantley et al., 2007). Our results in isolated islets from 5-week-old mice showed that, at this early stage, the amount of mRNA encoding IRS2 increased equally in *ob/ob* and POKO mice compared with WT and PPARγ2 KO mice. Similar results were obtained with staining for Ki67 (Fig. 1C), another marker of proliferation, which increased by a similar extent in β cells from *ob/ob* and POKO mice islets compared with WT and PPARγ2 KO controls. Despite these similar changes, we identified an important cell proliferation marker, cyclin D1, which is impaired selectively in the POKO mice. There is a greater increase in the expression of cyclin D1 in islets from *ob/ob* mice than in islets from POKO mice (although the expression in both of these genotypes is higher than in lean WT and PPARγ2 littermates) (Fig. 1B). Together, our results indicate that, compared with WT and PPARγ2 KO mice, proliferative programmes are activated in both *ob/ob* and POKO islets, even at this early stage, in response to insulin resistance. Furthermore, our results also highlight selective defects in cell proliferative genes in islets from POKO mice, which might indicate a subsequent failure of proliferation. Of interest, gene expression analysis in 5-week-old C57BL/6 female mice also revealed decreased levels of PPARγ1 in isolated islets from PPARγ2 KO and POKO mice, whereas the level of mRNA encoding PPARγ1 was maintained in *ob/ob* islets (Fig. 1B). From this, we speculated that the differences between *ob/ob* and POKO mice might be apparent in processes that are regulated by PPARγ (see below).

Markers of β-cell function that are associated with adaptation to insulin resistance are not apparent in islets from POKO mice

We undertook a systematic evaluation of the steps involved in coupling glucose concentration to insulin secretion. First, we investigated whether markers of glucose transport into islets are altered in POKO islets compared with the islets of other genotypes studied. Glut2 is the major glucose transporter involved in glucose uptake in islets, and is under the regulatory control of the pancreatic and duodenal homeobox 1 transcription factor (PDX1) (Lebrun et al., 2005). Glut2 is considered to be a surrogate of the state of differentiation of β cells. In agreement with this, we found that the expression levels of the mRNAs encoding both Glut2 and PDX1 are higher in islets from *ob/ob* mice, but that these increases were not present in POKO mice, which show similar expression levels to WT and PPARγ2 KO mice (Fig. 2A).

Next, we investigated the expression of genes involved in glucose intermediary metabolism, which is thought to be essential for

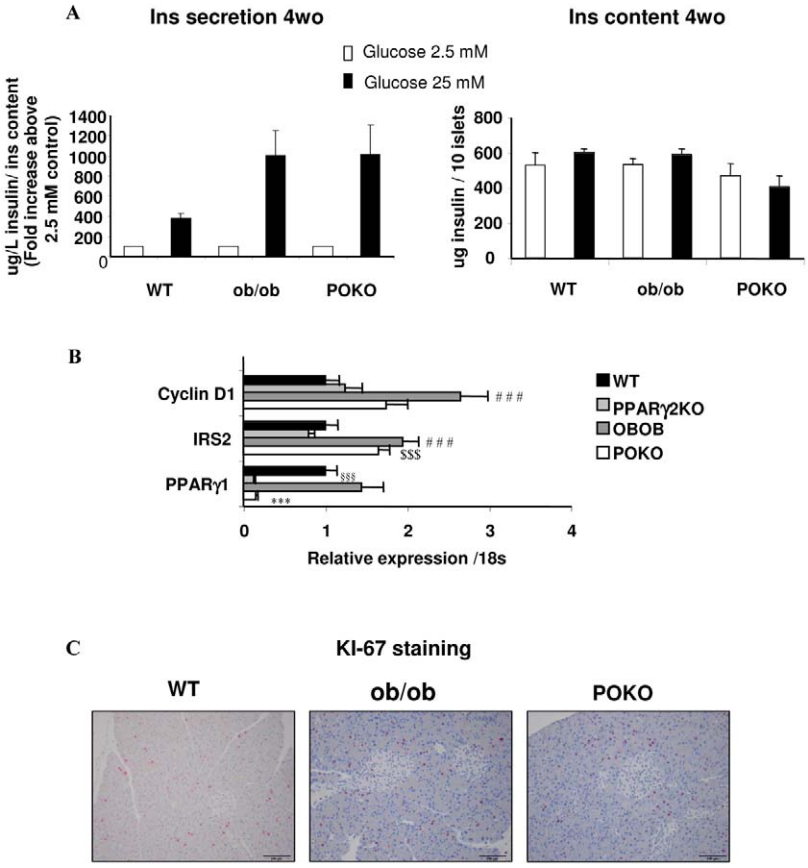


Fig. 1. β-cell function and proliferation in *ob/ob* and POKO mice. (A) Insulin secretion from islets isolated from WT, *ob/ob* and POKO mice in response to either 2.5 mM (white bars) or 25 mM (black bars) of glucose. Triplicate samples of ten different islets were obtained from each mouse (n=3-6 mice per genotype). For each sample, insulin release was normalised to insulin content and the fold increase in insulin above the 2.5 mM control mice was recorded. (B) Islet mRNA levels from different genes from 5-week-old female WT, PPARγ2 KO, *ob/ob* and POKO mice (n=8-11 mice per genotype). $^{###}P$<0.001 WT vs *ob/ob*; $^{§§§}P$<0.001 PPARγ2 KO vs WT; $^{$$$}P$<0.001 POKO vs PPARγ2 KO; $^{***}P$<0.001 POKO vs *ob/ob*. (C) Immunohistochemical analysis of Ki67 in the pancreas from 4-week-old male WT, *ob/ob* and POKO mice (n=5). Bars, 100 μm.

coupling glucose sensing to insulin release in β cells (Chen et al., 1994a; Chen et al., 1994b; Liu et al., 2002). An interesting difference between both young *ob/ob* and POKO mice was seen in the expression of mRNA encoding glucokinase (GK), which is upregulated by PPARγ agonists in β cells (Kim et al., 2002). In our studies, GK mRNA is increased in islets from *ob/ob* mice but not from POKO mice, in which expression levels remained comparable to WT and PPARγ2 KO mice (Fig. 2A). Expression of the gene pyruvate carboxylase (PCX), which is another target gene of PPARγ (Jitrapakdee et al., 2005), is decreased in PPARγ2 KO islets. As expected, the level of mRNA encoding PCX is also significantly decreased in islets from POKO mice compared with *ob/ob* mice (Hasan et al., 2008).

Compensation for insulin resistance requires a concomitant increase in insulin biosynthesis (Melloul et al., 2002). Our results indicate that, at the age of 4 weeks, both *ob/ob* and POKO mice can increase insulin gene expression to a level that is similar to WT mice. However, despite preserved expression of the insulin gene, transcription factors involved in insulin biosynthesis, such as v-maf musculoaponeurotic fibrosarcoma oncogene homologue A (MafA) (Olson et al., 1993; Hagman et al., 2005), are reduced in islets from PPARγ2 KO and POKO mice compared with WT and *ob/ob* mice (Fig. 2A).

Dysregulation of fatty acid metabolism in islets in young POKO mice

The mechanisms by which free fatty acids (FFAs) amplify glucose-induced insulin secretion have not been elucidated fully. Nevertheless, increased cytosolic malonyl-coenzyme A (CoA) that arises from glucose and de novo lipogenesis can act through AMP-activated protein kinase (AMPK)/malonyl-CoA pathways to limit fatty acid oxidation (Roduit et al., 2004). Interestingly, at these early stages, the levels of mRNA encoding sterol regulatory element-binding protein 1c (SREBP1c) are already increased in islets from *ob/ob* mice compared with WT, PPARγ2 KO and POKO mice. Of interest, POKO mice failed to show an increase in SREBP1c mRNA levels compared with *ob/ob* islets (Fig. 2B). However, at this stage, there were still no differences between *ob/ob* and POKO islets in their levels of expression of the mRNAs encoding acetyl-CoA carboxylase (ACC, also known as ACACA) or carnitine palmitoyltransferase 1 (CPT1) (Fig. 2B).

We have shown previously that the adult 16-week-old POKO islets contain a higher concentration of ceramides than *ob/ob* islets (Medina-Gomez et al., 2007). Here, we investigated whether defects in lipid metabolism were already present in islets at 4-5 weeks of age. Although we observed a trend, there were no statistically significant changes in lipid composition among the four genotypes of mice, as determined by analysis of variance (ANOVA) that was adjusted for multiple hypotheses (Fig. 3A). The levels of expression of mRNA encoding acid sphingomyelinase (ASM, SMPD1), an enzyme involved in generating ceramides from sphingomyelin, were increased in islets from both *ob/ob* and POKO mice compared with WT mice, although expression was higher in the *ob/ob* mice than in the POKO mice (Fig. 2B). Also, there were no significant changes in de novo ceramide synthesis

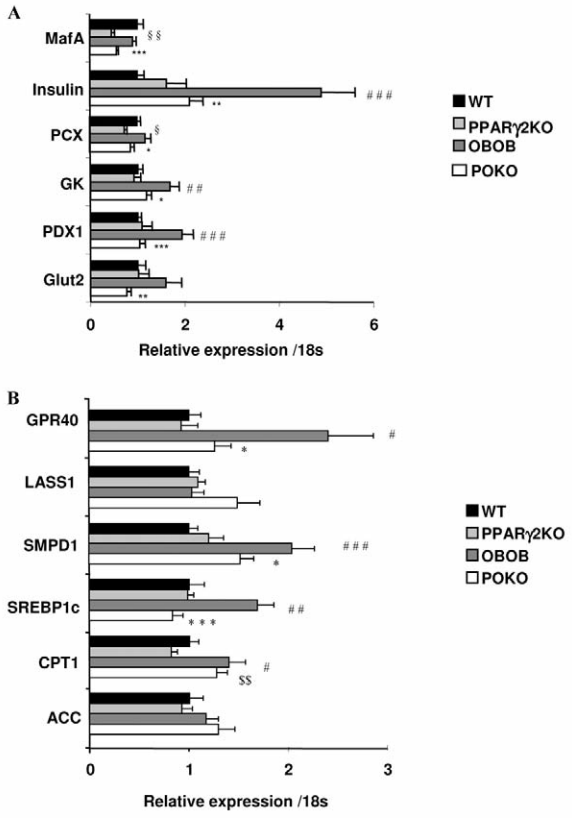


Fig. 2. Islet gene expression of glucose and lipid metabolism genes. (A) Islet gene expression from 5-week-old female WT, *ob/ob*, PPARγ2 KO and POKO mice (*n*=8-11 mice per genotype). (B) Islet gene expression of lipid metabolism genes from 5-week-old female WT, *ob/ob*, PPARγ2 KO and POKO mice (*n*=8-11 mice per genotype). $^{\#}P<0.05$, $^{\#\#}P<0.01$, $^{\#\#\#}P<0.001$ WT vs *ob/ob*; $^{\S}P<0.05$, $^{\S\S}P<0.01$ PPARγ2 KO vs WT; $^{\S\S}P<0.01$ POKO vs PPARγ2 KO; $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$ POKO vs *ob/ob*.

among the four genotypes, although there was a trend towards higher expression of ceramide synthase 1 (LASS1) in POKO islets compared with WT, PPARγ2 KO and *ob/ob* islets, which is in line with the increased levels of ceramides that were detected in older mice. Furthermore, expression of the gene encoding the G-protein-coupled receptor GPR40, a membrane receptor activated by lipids and involved in lipotoxicity (Itoh et al., 2003; Itoh and Hinuma, 2005), was higher in islets from *ob/ob* mice compared with WT, PPARγ2 KO and POKO mice (Fig. 2B). Our data indicate that, at this early age, islets from *ob/ob* and POKO mice are relatively protected from the accumulation of reactive toxic lipid species.

Early lipidomic changes in metabolically relevant organs in young POKO mice

Next, we investigated whether the relatively preserved lipid composition in POKO islets can also be observed in other metabolically relevant organs. In contrast to β cells, a multivariate analysis revealed large differences in the lipid profiles of serum, liver, adipose tissue and muscle (see below) in WT, PPARγ2 KO, *ob/ob* and POKO mice at the age of 4 weeks.

As expected from the hypertriglyceridaemia that was observed at 4 weeks of age, the serum from POKO mice had high levels of short-, medium- and long-chain triacylglycerols (TAGs), and high levels of short- and medium-chain phosphatidylcholine (PC) lipid species compared with the other genotypes studied (Fig. 3B). Interestingly, the levels of long-chain TAGs in the serum from young *ob/ob* mice were lower than in WT and PPARγ2 KO mice.

Lipidomic analyses of the liver also revealed differences in lipid profiles at 4 weeks of age (Fig. 4), with an increase in short- and medium-chain diacylglycerols (DAGs) and TAGs in *ob/ob* and POKO livers compared with WT livers; there was also a slight increase in these lipids in PPARγ2 KO livers. We also observed enrichment in unsaturated long-chain TAGs in POKO and *ob/ob* livers. Polyunsaturated long-chain TAGs remained at similar low levels in both POKO and *ob/ob* livers when compared with WT and PPARγ2 KO mice (Fig. 4). Although the ceramide concentration was similar in PPARγ2 KO, POKO and *ob/ob* livers, it was higher than in WT mice at this early age. The levels of medium-chain PCs also increased in livers from *ob/ob* and POKO mice compared with WT and PPARγ2 KO mice.

With respect to adipose tissue, at this stage, POKO and *ob/ob* mice had similar levels of short-, medium- and long-chain TAGs (see supplementary dataset). Unlike in the liver, the concentration of polyunsaturated long-chain TAGs with a high number of double bonds was higher in adipose tissue from POKO and *ob/ob* mice than from WT and PPARγ2 KO mice. The ceramide concentration was also increased to similar levels in the POKO and *ob/ob* genotypes.

Lipidomic analysis of muscle showed an increased concentration of short- and medium-chain TAGs in *ob/ob* mice compared with the other genotypes (Fig. 5). Conversely, muscle from POKO and PPARγ2 KO mice contained more long-chain TAGs than WT and *ob/ob* muscles. Compared with WT and PPARγ2 KO mice, muscles from *ob/ob* and POKO mice were enriched in ceramides, which correlated with the extent of the increased levels of lysoPCs.

Oxidative stress is not an early mechanism linked to β-cell failure in POKO mice

The toxic effects of increased ROS are proposed to mediate β-cell failure (Maedler et al., 2002; Kaneto et al., 2005). Here, we investigated whether the excessive toxicity of ROS is an early pathogenic mechanism that determines the divergence in β-cell phenotype between *ob/ob* and POKO mice. The production of ROS, which was measured using CM-H_2DCFDA, a cell-permeable indicator of ROS, revealed no differences between POKO and *ob/ob* islets (Fig. 6A). Uncoupling protein 2 (UCP2) is a mitochondrial carrier that is activated by ROS and reduces insulin secretion. We observed no differences in the levels of mRNA encoding UCP2 in islets, which supports the notion that ROS do not contribute to the early stages of β-cell failure in the POKO mouse (supplementary material Fig. S1A). Furthermore, the expression of mRNA encoding ROS scavengers that are typically induced during oxidative stress, such as manganese superoxide dismutase (MnSOD), is unchanged in islets from POKO and *ob/ob* mice compared with WT mice. The expression of other intracellular antioxidant enzymes such as glutathione peroxidase 1 (GPX-1) and glutathione reductase 1 (GSR-1) were also similar in the four genotypes, which further indicates that, at this age, differences in oxidative stress are unlikely

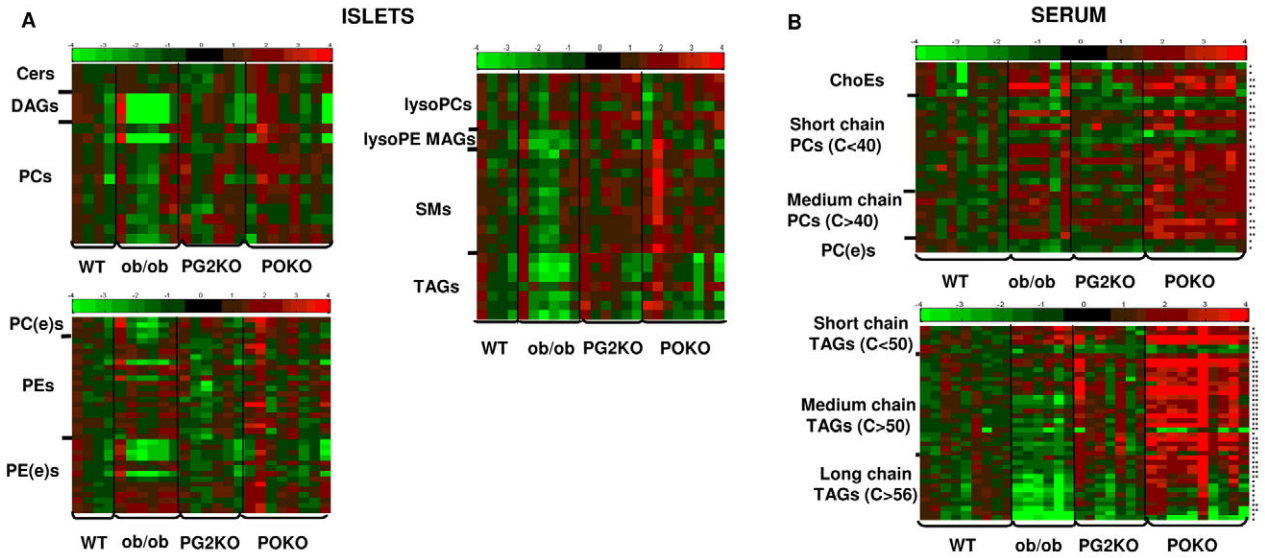


Fig. 3. Lipidomic profiling of islets and serum. Lipidomic profiling of islets (A) and serum (B) from 4-week-old male WT, PPARγ2 KO (PG2KO), *ob/ob* and POKO mice (n=5-8 mice per genotype). Lipids with ANOVA P values of $P<0.05$ are shown. Lipid variable P values: $^{*}P<0.05$, $^{**}P<0.01$. Abbreviations: Cers, ceramides; ChoEs, cholesteryl esters; DAGs, diacylglycerols; lysoPCs, lysophosphatidylcholines; lysoPE MAGs, lysophosphatidylethanolamine monoacylglycerols; PCs, phosphatidylcholines; PC(e)s, ether-linked phosphatidylcholines; PEs, phosphatidylethanolamines; PE(e)s, ether-linked phosphatidylethanolamines; SMs, sphingomyelins; TAGs, triacylglycerols.

to be a main determinant of the rapid evolution of β-cell failure in pancreatic islets of the POKO mice.

Macrophage infiltration, islet inflammation and fibrosis in POKO islets

Next, we investigated whether inflammation was present in islets at this early stage and, therefore, whether this might be an early pathogenic mechanism responsible for the accelerated β-cell failure in islets of POKO mice. At 4 weeks of age, the expression of the macrophage markers CD68 and F4/80 is increased in islets from *ob/ob* and POKO mice compared with WT and PPARγ2 KO mice, indicating that macrophage infiltration in islets is an early event in the development of the metabolic syndrome. However, at this early age, macrophage infiltration of islets per se does not predict the subsequent evolution towards β-cell failure in POKO mice. In support of this, we observed no differences in the expression of interleukin (IL)-6 and tumour necrosis factor (TNF)α mRNA in islets (Fig. 6B) and/or in the level of fibrosis, as assessed by Picrosirius Red staining in 4-week-old mice from the four genotypes (supplementary material Fig. S2). Similarly, there was no evidence of increased apoptosis in islets using a TUNEL assay (Fig. 6C), and no difference in the expression of Bcl2 in POKO, *ob/ob*, WT and PPARγ2 KO mice (Fig. 6B). Together, these data indicate that although *ob/ob* and POKO mice show some proinflammatory changes at 4 weeks of age, these changes do not predict the future divergence in β-cell failure in these mice.

ER stress is not an early pathogenic mechanism leading to β-cell failure in POKO mice

Metabolically overstretched β cells make high demands on the ER for the biosynthesis of insulin (Nakatani et al., 2005). As these demands become chronic, the biosynthesis of insulin might overload the protein-folding capacity of the ER and so play a role in the development of β-cell failure (Kaneto et al., 2005; Nakatani et al., 2005; Wellen and Hotamisligil, 2005). We investigated whether early differences in ER stress markers could predict the subsequent evolution of β-cell function in *ob/ob* and POKO mice. Initially, we assessed X-box binding protein 1 (XBP1), a transcription factor that is downstream of inositol-requiring protein 1 (IRE1) and that is a transmembrane protein in the ER, which functions as a sensor and transducer of ER stress (Eizirik et al., 2008). The mRNA levels of the precursor form of XBP1 (unspliced) and the active (spliced) form, which is formed by an IRE1-mediated splicing reaction following ER stress, were similar in the four genotypes (supplementary material Fig. S1B). Moreover, an ER overload leads to apoptosis through the induction of the C/EBP homologous protein Chop (Huang et al., 2007; Laybutt et al., 2007; Marchetti et al., 2007). Although the mRNA expression level of Chop is increased in both *ob/ob* and POKO mice at 5 weeks of age compared with WT and PPARγ2 KO mice, this increase is lower in POKO mice than in *ob/ob* mice. However, expression of the ER chaperone glucose-regulated protein 78 (GRP78) is similar in *ob/ob* and POKO mice compared with WT and PPARγ2 KO mice (supplementary material Fig. S1B). This result indicates that although incipient ER stress occurs in the early stages of β-cell compensation, it alone cannot be used to predict the course of development of β-cell expansion and failure.

DISCUSSION

Our goal was to characterise the early events and molecular factors leading to β-cell adaptation and failure in the context of obesity and insulin resistance. For this, we took advantage of a recently generated mouse model, the POKO mouse, which was obtained by crossing a PPARγ2 KO mouse with an *ob/ob* mouse, as described previously (Medina-Gomez et al., 2007). This study focuses on the early events that lead to severe metabolic syndrome. Our hypothesis

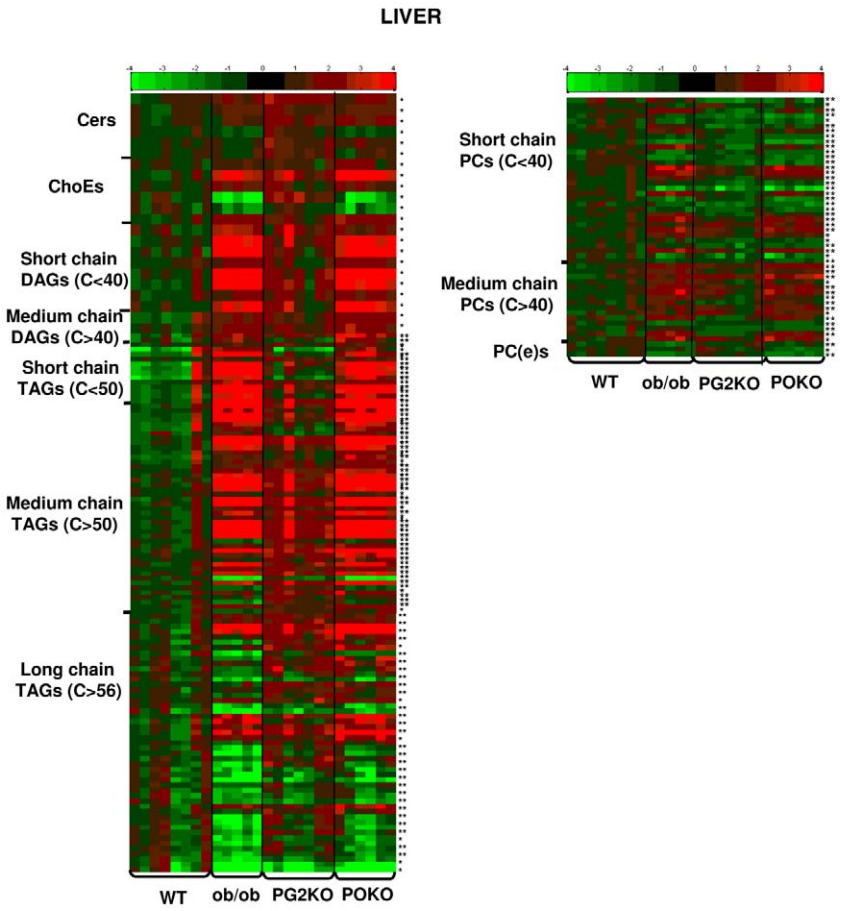


Fig. 4. Lipidomic profiling of the liver. Lipidomic profiling of the liver from 4-week-old male WT, PPARγ2 KO (PG2KO), *ob/ob* and POKO mice (*n*=5-8 mice per genotype). Lipids with ANOVA *P* values of $P<0.05$ are shown. Lipid variable *P* values: $*P<0.05$, $**P<0.01$.

is that the early stages of disease might provide important novel clues in the pathogenesis of metabolically induced β-cell failure and diabetes.

Here, we show that differences in metabolic stress are detected as early as 4-5 weeks of age, as illustrated by the differences observed between *ob/ob* and POKO islets at a time when the mass and functional qualities of their β cells are apparently conserved. Previous studies addressing this subject have faced the problem of discriminating primary pathogenic mechanisms from secondary toxic hits, or even from protective secondary allostatic responses that maintain the metabolic homeostasis of the system. Here, we focused on early events in an attempt to unravel the differences that, at an early age, could be predictive of β-cell failure. The strength of our experimental design is based on comparing two mouse models that have different degrees of severity and kinetics of β-cell failure and on using PPARγ2 KO and WT mice, which are not insulin resistant and have normal β-cell function, as controls. Our data indicate that, despite their apparently normal β-cell morphology and function, metabolic and genetic differences in young *ob/ob* and POKO mice differentially affect β-cell function and lipid metabolism in response to insulin resistance. We observed that other well-established mechanisms that are associated with β-cell failure, such as increased ROS production, inflammation, macrophage infiltration and ER stress, are already present at this early age. As these proinflammatory mechanisms are equally apparent in POKO and *ob/ob* mice, it is possible that the main difference between β-cell failure in POKO and *ob/ob* mice might be related to decreased resistance to stress. Our data also indicate that these measures cannot be used to predict β-cell failure, at least in our models.

To validate the POKO mouse as a model of β-cell failure, and to validate its potential use in translational studies, we initially characterised specific candidate mechanisms that have been suggested previously to be involved in β-cell failure in humans. As indicated above, our results represent an in-depth characterisation of the early metabolic events in β cells in two mouse models that both eventually develop metabolic syndrome, although with differing severity. As such, their comparative analysis offers the opportunity of identifying predictors of the metabolic syndrome development and severity. One model, the POKO mouse, develops severe metabolic syndrome owing to positive energy balance, impaired adipose tissue expandability and accelerated β-cell failure (Medina-Gomez et al., 2007). The *ob/ob* mouse is a genetic model of positive energy balance that results from increased food intake and decreased energy expenditure. Adipose tissue expansion is not impaired in this model; nevertheless, the mice develop the metabolic syndrome. One of the most puzzling, and interesting, findings of the initial characterisation of adult POKO mice was their inability to expand β-cell mass, a response that is typical of other insulin-resistant strains, such as their *ob/ob* littermates. Another interesting observation is that, despite the marked differences in β-cell mass

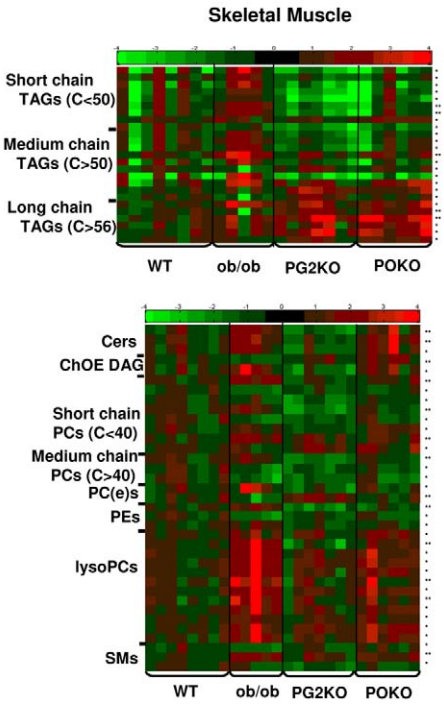


Fig. 5. Lipidomic profiling of skeletal muscle. Lipidomic profiling of skeletal muscle from 4-week-old male WT, PPARγ2 KO (PG2KO), *ob/ob* and POKO mice (*n*=5-8 mice per genotype). Lipids with ANOVA *P* values of $P<0.05$ are shown. Lipid variable *P* values: $*P<0.05$, $**P<0.01$.

in adult mice, β-cell mass and function in younger (4-5 weeks old) POKO and *ob/ob* mice are indistinguishable, which indicates that this might be an interesting time at which to explore early mediators of the subsequent divergent response.

Our results confirm that, at 4 weeks of age, POKO and *ob/ob* mice have a similar β-cell mass (Medina-Gomez et al., 2007), and conserved basal and stimulated insulin secretion. Conservation of the insulin secretory capacity of *ob/ob* and POKO islets indicates that, at this stage, islets from these models are relatively healthy. This led us to speculate that the differences observed at this stage should be enriched in primary adaptations in comparison to the more likely secondary adaptive and/or toxic responses that become apparent in advances stages of the disease.

Our results confirm that it is possible to work efficiently with such small islets and to detect signs of metabolic stress in this tissue. We observed upregulation of IRS2 mRNA, a key mediator of β-cell proliferation that protects β cells against glucotoxicity. We also observed upregulation of mRNA encoding Ki67, which is another marker of cell proliferation that stains from the early to mid-G1 phase of the cell cycle through to mitosis. Given the adaptive increase in β-cell mass to maintain euglycaemia in the presence of marked insulin resistance, the increased expression of cyclin D1 in *ob/ob* islets was not unexpected. Although this may appear paradoxical in leptin-deficient mice, as cyclin D1 has been shown to be a target of leptin (Saxena et al., 2007), it indicates that other specific regulators of cyclin D1 regulate β-cell mass expansion. Our data also indicate that induction of cyclin D1 in β cells might be regulated, in part, by PPARγ under conditions of either overnutrition or metabolic stress. However, this hypothesis is challenged by the observation that cyclin D1 is not altered in islets from PPARγ2 KO mice, even though pharmacological activation of PPARγ decreases cyclin D1 expression in vitro (Yin et al., 2001; Koga et al., 2003). Alternatively, it has been suggested that agonist-mediated inhibition of cell growth might be independent of PPARγ activation (Bae et al., 2003). Despite the opposing roles of leptin and PPARγ in modulating β-cell mass, our data indicate that robust alternative mechanisms are able to expand β-cell mass appropriately under conditions of increased metabolic demands. These studies also lead us to speculate that a systems biology approach to studying POKO islets might be a useful way to identify these mechanisms. In any case, defining the role of cyclin D1 in coupling β-cell expansion to insulin resistance is an important subject for subsequent studies.

Our data are sufficiently sensitive to detect specific metabolic disturbances in islets from young mice. For example, Glut2, an important transporter required for glucose uptake in islets, and its main upstream regulator PDX1, are induced in islets from *ob/ob* mice. In POKO islets, however, expression of Glut2 and PDX1 is similar to that observed in WT islets, which could be considered similar to unstressed physiological levels. In addition to compromised glucose uptake into islets from POKO mice, the early steps of glucose metabolism are likely to be affected, as indicated by selective defects in GK and PCX expression in islets (Hasan et al., 2008). As suggested above, these genes appear to be bona fide PPARγ targets because GK is regulated by thiazolidinediones in β cells (Kim et al., 2002), and our group has evidence that PCX is a target of PPARγ action (Jitrapakdee et al., 2005). This reinforces the idea that the POKO model might be enriched in defects in PPARγ-dependent mechanisms of β-cell homeostasis. Similarly, these data indicate that PPARγ targets might be important for the adaptation of β cells to insulin resistance and nutritional stress. As expected, expression of the gene encoding insulin is increased in islets from 5-week-old *ob/ob* and POKO mice compared with non-insulin-resistant PPARγ2 KO and WT islets. However, it might be argued that the level of upregulation in POKO mice is inappropriately low considering that these mice are more insulin resistant than *ob/ob* mice. Gene expression analysis further supports the concept of inappropriate adaptation to metabolic stress in POKO mice, as indicated by the blunted transcriptional upregulation of MafA and PDX1 in POKO islets compared with the marked upregulation that was observed in *ob/ob* islets. Our studies of glucose sensing and fuel handling also distinguish between the severity of metabolic stress in the two models. When considered globally, islets from POKO mice showed inappropriate adaptations in β-cell function, and defects in proliferation markers and lipid signalling. Interestingly, compared with WT islets, the expression of metabolic genes in POKO islets was unchanged, with a profile that could be considered to be inappropriately normal for the severe metabolic stress.

Following our original data from adult POKO mice (Medina-Gomez et al., 2007), we sought to identify whether defects in fatty acid metabolism and lipotoxicity were already evident in islets from young POKO mice. The importance of the role of fatty acid metabolism in β-cell function is illustrated by the fact that FFAs amplify glucose-induced insulin secretion, although the mechanisms of this effect are far from being elucidated. It has

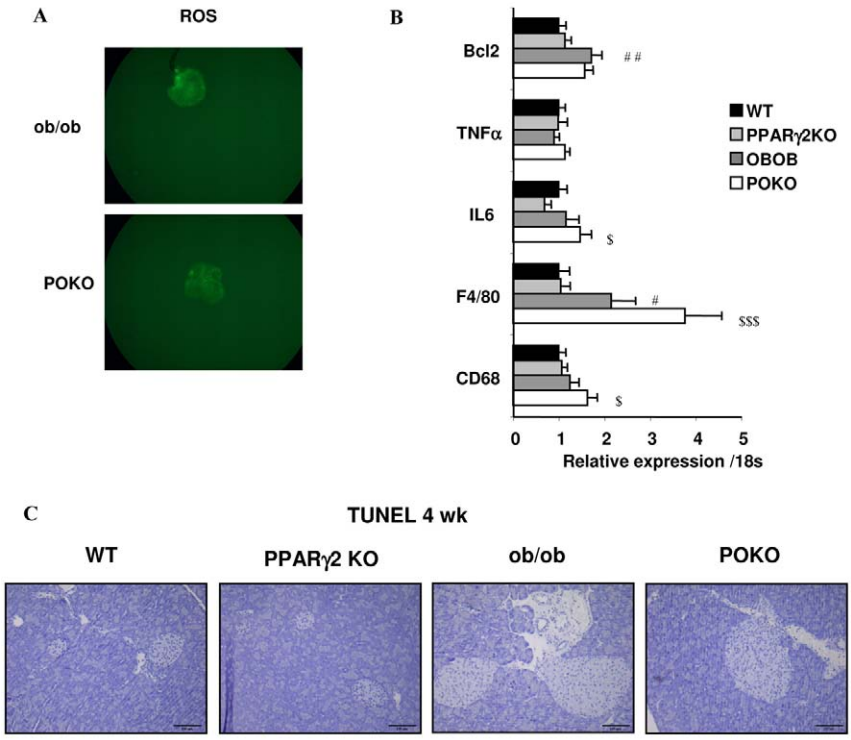


Fig. 6. Islet ROS production, gene expression and apoptosis in young mice. (A) ROS production was measured by CM-H_2DCFDA fluorescence, as described in the Methods. (B) Islet gene expression from 5-week-old female WT, *ob/ob*, PPARγ2 KO and POKO mice (n=8-11 mice per genotype). $^{\#}P<0.05$, $^{\#\#}P<0.01$ WT vs *ob/ob*; $^{\$}P<0.05$, $^{\$\$\$}P<0.001$ POKO vs PPARγ2 KO.
(C) Representative pancreatic sections showing apoptosis in WT, *ob/ob*, PPARγ2 KO and POKO mice, determined by TUNEL. Bars, 100 μm.

been suggested that inappropriate activation of AMPK/ACC/malonyl-CoA/CPT1 pathways might limit fatty acid oxidation and compromise β-cell function (Roduit et al., 2004). In young mice, there were no differences in the expression of ACC or CPT1 mRNA. However, even at this early stage, expression of the gene encoding SREBP1c was decreased in POKO islets compared with *ob/ob* islets. This is an interesting observation because there is evidence that excessive activation of SREBP1c reduces insulin secretion and genetically modified mouse models have indicated that SREBP1c has a physiological role in β cells (Yamashita et al., 2004; Takahashi et al., 2005; Shimano et al., 2007). Furthermore, recent evidence indicates that enhanced lipid synthesis that is mediated by SREBP1c-dependent genes might be required for the adaptive changes in islet gene expression and insulin secretion that occur at high glucose concentrations (Diraison et al., 2008). Induction of SREBP1c might, in our opinion, contribute to the hormetic effect of fatty acids on insulin secretion. In this regard, the increase in SREBP1c that was observed in *ob/ob* mice, compared with WT mice, might be an adaptation to promote lipogenesis in response to glucose levels; this adaptation seems to be missing in islets from POKO mice, despite similar challenging metabolic conditions.

We have also optimised a sensitive and sophisticated lipidomic platform that is sufficiently sensitive to characterise the small number of islets isolated from 4-week-old mice. Unlike in adult mice (Medina-Gomez et al., 2007), the lipid composition of islets from young POKO mice was unchanged when compared with the other genotypes. However, lipidomic analysis of other metabolically relevant tissues revealed important changes in lipid species in young animals. For example, livers from both POKO and *ob/ob* mice had similarly increased TAG levels, predominantly short- and medium-chain TAGs, at 4 weeks of age. A priori, we would expect POKO mice to accumulate more TAGs than *ob/ob* mice, particularly considering the decrease in adipose tissue expandability that is secondary to ablation of PPARγ2, and the severity of the metabolic disturbances (e.g. hyperlipidaemia) in POKO mice compared with the other genotypes. In this respect, the similarity between TAG accumulation in the livers of *ob/ob* and POKO mice is compatible with some degree of impaired TAG deposition in the POKO liver.

Another interesting finding comes from analysis of skeletal muscle from 4-week-old POKO mice; these data showed that the concentrations of short- and medium-chain TAGs increased by less than in skeletal muscle from *ob/ob* mice and were associated with an increased concentration of other lipid species, such as long-chain TAGs. This result agrees with the lipid profile observed in the POKO mouse at later stages. Of interest, the levels of ceramides and proinflammatory lysoPCs increased similarly in both *ob/ob* and POKO mice. Overall, our lipidomic studies suggest that the defective accumulation of TAGs and the resulting lipotoxicity in relevant metabolic tissues, coupled with accelerated β-cell failure and the relatively low levels of insulin, may contribute to the severity of the metabolic syndrome and the lipidomic profile observed in mature POKO mice. Our data also highlight qualitative and quantitative differences in organ-specific lipid networks and their contribution to insulin resistance. These lipidomic studies and, more specifically, the differences observed between the liver, serum and muscle, may also indicate the possibility of a hierarchical order of organs with respect to fat deposition and lipid-induced toxicity. In this respect,

our results indicate that, when adipose tissue storage becomes incompetent, islets are not the first tissue to be to be affected by lipid insult; it is more likely that the liver is the first organ to serve as a sink for fat.

Our analysis of the early mechanisms leading to β-cell adaptation and failure provide evidence that islet inflammation, ROS and ER stress, which are all alterations that are observed in the context of severe advanced apoptosis of β cells (Poitout and Robertson, 2002), are already present in islets from young POKO and *ob/ob* mice but not in islets from PPARγ2 KO and WT mice. However, we do not conclude that these proinflammatory mechanisms are predictive markers of either severity or the prognosis of β-cell failure. In our opinion, it is more likely that the primary mechanism, in this case the ablation of PPARγ, defines the vulnerability of the β cells to these pathogenic mechanisms.

In summary, we have characterised the early metabolic alterations in islets from *ob/ob* mice and compared them with islets of the more severely affected POKO mice, and then compared these to the changes in WT and PPARγ2 KO controls. Our results confirm that β cell studies at this early age are both feasible and sufficiently sensitive to identify altered processes, and to discriminate between different degrees of severity. Our results justify future systems biology studies to investigate the early stages of β-cell disease as a strategy to maximise the chances of identifying primary pathogenic mechanisms over secondary hits and allostatic changes. Our lipidomic approach reveals that the lipid profile of metabolically relevant organs changes during the course of the disease, and that these changes are specific to the particular organ. Taken together, our data support the concept that studies in young mouse models are informative; that *ob/ob* mice, and particularly the accelerated phenotype of POKO mice, provide good models to study the early mechanisms associated with β-cell failure; and that coupling studies in β cells with lipid analysis in other organs might provide important information about the chronology of lipid disturbances in specific organs in the context of lipotoxicity and the metabolic syndrome.

METHODS

Animal care

Animals were housed at a density of four animals per cage in a temperature-controlled room (20-22°C) with 12-hour light-dark cycles. Food and water were available ad libitum unless noted. All animal protocols used in this study were approved by the UK Home Office and the University of Cambridge.

Isolation and culture of pancreatic islets

The pancreas was injected, though the bile duct, with cold Hank's solution containing 0.4% (w/v) collagenase P (Roche Biochemicals). The pancreas was removed, digested for 8 minutes, and islets were collected by hand picking. Isolated islets were cultured overnight in cell medium (RPMI 1640 with 10% FBS, and penicillin and streptomycin), at 37°C, in 5% CO_2 in air. For insulin secretion studies or RNA extraction, islets were used on the day after isolation.

Insulin secretion studies

Insulin secretion from isolated islets (10 islets/well) was measured for 30 minutes at 37°C in incubations in Krebs-Ringer buffer, supplemented with 0.1% bovine serum albumin (BSA) as a carrier containing either basal (2.5 mM) or stimulatory (25 mM) glucose concentrations. The supernatants were assayed for insulin. Insulin content was extracted using a 95:5 ethanol:acetic acid solution. Insulin was measured using an electrochemical luminescence immunoassay from MesoScale Discovery (MSD) (Gaithersburg, MD). For these experiments, islets were isolated from several mice of each genotype. Thus, the data are the mean of separate experiments in which data were collected for each test solution from samples of 10 islets each. For each sample, insulin release was normalised to insulin content.

RNA preparation and real-time quantitative PCR

Total RNA was isolated from islets and tissue samples according to the manufacturer's instructions (RNAeasy kit, Qiagen). Real-time quantitative PCR was performed using a TaqMan 7900 (Abi), according to standard protocols.

Light microscopy and immunohistochemical analysis

Tissue samples for morphological and immunohistochemical analyses were prepared according to published protocols (Medina-Gomez et al., 2007). Morphometric analyses of pancreas sections were acquired using a digital camera and microscope (Olympus BX41), and cell areas were measured using AnalySIS software (Soft Imaging System). The antibodies used for immunohistochemistry were as follows: Ki67 staining was detected using streptavidin alkaline phosphatase (Zymed) at 1:2000 in TTBSA (Tris-Tyrode's buffer supplemented with BSA), and visualised with SigmaFast Fast Red TR/Naphthol AS-MX phosphate tablets. TUNEL staining slides were pretreated with Dako Real proteinase K for 5 minutes and then the staining was performed using Millipore's ApopTag peroxidase in situ apoptosis detection kit (Chemicon S7100), according to the manufacturer's instructions. For estimation of collagen deposition, slides were stained with Weigert's iron haematoxylin (made in-house) for 30 minutes, then briefly differentiated and counterstained with Picrosirius Red (Pioneer Research chemicals) for 1 hour, rapidly dehydrated, cleared, and mounted in Micromount.

Measurement of ROS production

To determine the production of ROS in whole islets, islets were loaded with 5 mol/l of CM-H_2DCFDA for 1 hour followed by three washes with fresh culture medium. ROS production was visualised on a fluorescent microscope equipped with a digital camera. One representative figure from three independent experiments with similar results is shown (Fig. 6A).

Lipid profiling

Serum samples (10 μl), tissue samples (about 20 mg for liver, white adipose tissue and skeletal muscle tissue) or islet samples were diluted with 0.9% NaCl (10 μl for serum and islets, 50 μl for tissue samples) and spiked with an internal standard reference compounds mixture (10 μl for serum and islets, 20 μl for tissue samples) (Laaksonen et al., 2006). The samples were subsequently extracted with chloroform:methanol (2:1) solvent (100 μl for serum, 90 μl for islets, 200 μl for tissue samples), homogenised with a glass rod, vortexed (1 minute for serum, 15 seconds for islets, 2 minutes for tissue samples), incubated at room temperature (1 hour for serum and tissue samples, 20 minutes for islets), and centrifuged at 7826 ***g***

TRANSLATIONAL IMPACT

Clinical issue

The earliest stages of type 2 diabetes are characterised by growing insulin resistance and increased insulin secretion from an expanding population of pancreatic β cells. In later stages, the β cells fail and, in the insulin-resistant environment, they cannot produce sufficient insulin to maintain blood glucose control. The ability to understand the crucial steps that precipitate diabetes and β-cell failure is hampered by limited mouse models for the early stages of the disease. Most models are examined after the failing of β cells, making it difficult to define the events leading to expansion and excessive insulin production by β cells, and their subsequent failure.

Results

Here, the authors define a mouse model for studying the early pathological changes that result from elevated fat levels or lipotoxicity. They use the POKO mouse, a mouse model obtained by crossing the insulin-resistant *ob/ob* mouse (which lacks leptin) with the peroxisome proliferator-activated receptor gamma 2 (PPARγ2) knockout mouse that lacks the nuclear receptor PPARγ2, which is known to have a role in adipogenesis. The POKO mouse becomes lipotoxic and more insulin resistant than the *ob/ob* mouse. Changes are found in the lipid profiles of POKO mice as young as 4 weeks of age and are associated already with changes in β-cell proliferation and function. At this early time, pathogenic effectors that characterise the advanced stages of disease can be observed. Lipidomic analysis of the β cells in pre-diabetic mice reveal that, compared with other metabolically relevant organs such as the liver, muscle or adipose tissue, these pancreatic cells are protected from lipid-induced toxicity at the early stages of disease. This suggests a hierarchical order of ectopic lipid accumulation in which β cells are protected initially, and establishes a model for understanding the early events leading to diabetes.

Implications and future directions

Previous models are limited in their ability to address issues about the early sensing of insulin resistance and the effectors that contribute to changes in β-cell mass and function. The authors suggest that these data establish a foundation for subsequent systems biology profiling studies to identify the mechanisms by which β cells initially sense insulin resistance and eventually fail from metabolic pressure. Further work should reveal the molecular effectors that increase β-cell mass and function in the early phase of diabetes and suggest biomarkers for early stages of the disease.

doi:10.1242/dmm.004499

for 3 minutes. From the separated lower phase, an aliquot was mixed with 10 ml of a labelled standard mixture (three stable isotope-labelled reference compounds), and a 1.0-ml injection was used for liquid chromatography-mass spectrometry (LC-MS) analysis. The sample order for analysis was established by randomisation. Lipid extracts were analyzed on a Q-ToF Premier mass spectrometer (Waters) combined with Acquity ultra-performance liquid chromatography (UPLC-MS) (Medina-Gomez et al., 2007) in ESI+ mode. Data processing was performed using the MZmine software (Katajamaa et al., 2006). The normalisation procedure for lipidomic data was as described previously (Laaksonen et al., 2006). Serum lipidomic data were normalised against the serum volume of samples, and lipidomic data from tissues were normalised against their tissue weights. Identification of lipid species was performed based on an internal database of a lipid library or, alternatively, by utilising the tandem mass spectrometry library.

Statistics

For lipidomic studies, an array view of lipid profiles reflected the changes relative to the mean intensity of molecular species within the WT group. One-way ANOVA was performed to investigate whether the mean intensities differed among the four genotypes. Lipids with ANOVA P values of $P<0.05$ were shown. The lipid variables with P values lying between $P<0.05$ and $P<0.01$ were marked with '*' and those with P values of $P<0.01$ were marked with '**'. The multivariate statistical data analyses were performed using Matlab (Mathworks) and the Matlab library PLS Toolbox (Eigenvector Research).

The rest of the results are given as the mean±S.E. (standard error). Statistical differences and interactions were evaluated through a two-way lack of leptin and lack of PPARγ2 factorial ANOVA. When statistically significant differences resulted at the interaction level, the Student's t-test was carried out to compare the effects. The differences were considered to be statistically significant at $P<0.05$.

ACKNOWLEDGEMENTS

We thank Keith Burling, Janice Carter and Daniel Hart for their work. We thank Sophie Gough, Maria Adams and Sergio Rodriguez-Cuenca for their comments on the writing of this manuscript. We thank the funding bodies that have supported the research within our laboratory leading to this work: Diabetes UK, FP6 Hepadip, and the MRC career establishment award and MRC CORD. Manuel Ros is a recipient of a Salvador de Madariaga fellowship from MEC, Spain. Deposited in PMC for release after 6 months.

COMPETING INTERESTS

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

G.M.-G. and A.V.-P. conceived and designed the experiments; G.M.-G., V.V., L.Y., M.C., M.J.-L., M.B. and M.R. performed the experiments; G.M.-G., V.V., L.Y. and M.O. analyzed the data; G.M.-G. and A.V.-P. wrote the paper.

SUPPLEMENTARY MATERIAL

Supplementary material for this article is available at
http://dmm.biologists.org/lookup/suppl/doi:10.1242/dmm.003251/-/DC1

Received 24 March 2009; Accepted 8 July 2009.

REFERENCES

Bae, M. A., Rhee, H. and Song, B. J. (2003). Troglitazone but not rosiglitazone induces G1 cell cycle arrest and apoptosis in human and rat hepatoma cell lines. *Toxicol. Lett.* **139**, 67-75.

Bollheimer, L. C., Skelly, R. H., Chester, M. W., McGarry, J. D. and Rhodes, C. J. (1998). Chronic exposure to free fatty acid reduces pancreatic beta cell insulin content by increasing basal insulin secretion that is not compensated for by a corresponding increase in proinsulin biosynthesis translation. *J. Clin. Invest.* **101**, 1094-1101.

Brownlee, M. (2003). A radical explanation for glucose-induced beta cell dysfunction. *J. Clin. Invest.* **112**, 1788-1790.

Cantley, J., Choudhury, A. I., Asare-Anane, H., Selman, C., Lingard, S., Heffron, H., Herrera, P., Persaud, S. J. and Withers, D. J. (2007). Pancreatic deletion of insulin receptor substrate 2 reduces beta and alpha cell mass and impairs glucose homeostasis in mice. *Diabetologia* **50**, 1248-1256.

Chen, C., Bumbalo, L. and Leahy, J. L. (1994a). Increased catalytic activity of glucokinase in isolated islets from hyperinsulinemic rats. *Diabetes* **43**, 684-689.

Chen, C., Hosokawa, H., Bumbalo, L. M. and Leahy, J. L. (1994b). Mechanism of compensatory hyperinsulinemia in normoglycemic insulin-resistant spontaneously hypertensive rats. Augmented enzymatic activity of glucokinase in beta-cells. *J. Clin. Invest.* **94**, 399-404.

Choudhury, A. I., Heffron, H., Smith, M. A., Al-Qassab, H., Xu, A. W., Selman, C., Simmgen, M., Clements, M., Claret, M., Maccoll, G. et al. (2005). The role of insulin receptor substrate 2 in hypothalamic and beta cell function. *J. Clin. Invest.* **115**, 940-950.

Diraison, F., Ravier, M. A., Richards, S. K., Smith, R. M., Shimano, H. and Rutter, G. A. (2008). SREBP1 is required for the induction by glucose of pancreatic beta-cell genes involved in glucose sensing. *J. Lipid Res.* **49**, 814-822.

Eizirik, D. L., Cardozo, A. K. and Cnop, M. (2008). The role for endoplasmic reticulum stress in diabetes mellitus. *Endocr. Rev.* **29**, 42-61.

Hagman, D. K., Hays, L. B., Parazzoli, S. D. and Poitout, V. (2005). Palmitate inhibits insulin gene expression by altering PDX-1 nuclear localization and reducing MafA expression in isolated rat islets of Langerhans. *J. Biol. Chem.* **280**, 32413-32418.

Hasan, N. M., Longacre, M. J., Stoker, S. W., Boonsaen, T., Jitrapakdee, S., Kendrick, M. A., Wallace, J. C. and MacDonald, M. J. (2008). Impaired anaplerosis and insulin secretion in insulinoma cells caused by small interfering RNA-mediated suppression of pyruvate carboxylase. *J. Biol. Chem.* **283**, 28048-28059.

Huang, C. J., Lin, C. Y., Haataja, L., Gurlo, T., Butler, A. E., Rizza, R. A. and Butler, P. C. (2007). High expression rates of human islet amyloid polypeptide induce endoplasmic reticulum stress mediated beta-cell apoptosis, a characteristic of humans with type 2 but not type 1 diabetes. *Diabetes* **56**, 2016-2027.

Itoh, Y. and Hinuma, S. (2005). GPR40, a free fatty acid receptor on pancreatic beta cells, regulates insulin secretion. *Hepatol. Res.* **33**, 171-173.

Itoh, Y., Kawamata, Y., Harada, M., Kobayashi, M., Fujii, R., Fukusumi, S., Ogi, K., Hosoya, M., Tanaka, Y., Uejima, H. et al. (2003). Free fatty acids regulate insulin secretion from pancreatic beta cells through GPR40. *Nature* **422**, 173-176.

Jitrapakdee, S., Slawik, M., Medina-Gomez, G., Campbell, M., Wallace, J. C., Sethi, J. K., O'Rahilly, S. and Vidal-Puig, A. J. (2005). The peroxisome proliferator-activated receptor-gamma regulates murine pyruvate carboxylase gene expression in vivo and in vitro. *J. Biol. Chem.* **280**, 27466-27476.

Kaneto, H., Matsuoka, T. A., Nakatani, Y., Kawamori, D., Miyatsuka, T., Matsuhisa, M. and Yamasaki, Y. (2005). Oxidative stress, ER stress, and the JNK pathway in type 2 diabetes. *J. Mol. Med.* **83**, 429-439.

Katajamaa, M., Miettinen, J. and Oresic, M. (2006). MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* **22**, 634-636.

Kim, H. I., Cha, J. Y., Kim, S. Y., Kim, J. W., Roh, K. J., Seong, J. K., Lee, N. T., Choi, K. Y., Kim, K. S. and Ahn, Y. H. (2002). Peroxisomal proliferator-activated receptor-gamma upregulates glucokinase gene expression in beta-cells. *Diabetes* **51**, 676-685.

Koga, H., Harada, M., Ohtsubo, M., Shishido, S., Kumemura, H., Hanada, S., Taniguchi, E., Yamashita, K., Kumashiro, R., Ueno, T. et al. (2003). Troglitazone induces p27Kip1-associated cell-cycle arrest through down-regulating Skp2 in human hepatoma cells. *Hepatology* **37**, 1086-1096.

Laaksonen, R., Katajamaa, M., Paiva, H., Sysi-Aho, M., Saarinen, L., Junni, P., Lutjohann, D., Smet, J., Van Coster, R., Seppanen-Laakso, T. et al. (2006). A systems biology strategy reveals biological pathways and plasma biomarker candidates for potentially toxic statin-induced changes in muscle. *PLoS ONE* **1**, e97.

Laybutt, D. R., Preston, A. M., Akerfeldt, M. C., Kench, J. G., Busch, A. K., Biankin, A. V. and Biden, T. J. (2007). Endoplasmic reticulum stress contributes to beta cell apoptosis in type 2 diabetes. *Diabetologia* **50**, 752-763.

Lebrun, P., Montminy, M. R. and Van Obberghen, E. (2005). Regulation of the pancreatic duodenal homeobox-1 protein by DNA-dependent protein kinase. *J. Biol. Chem.* **280**, 38203-38210.

Li, X., Chen, H. and Epstein, P. N. (2006). Metallothionein and catalase sensitize to diabetes in nonobese diabetic mice: reactive oxygen species may have a protective role in pancreatic beta-cells. *Diabetes* **55**, 1592-1604.

Liu, Y. Q., Jetton, T. L. and Leahy, J. L. (2002). beta-cell adaptation to insulin resistance: increased pyruvate carboxylase and malate-pyruvate shuttle activity in islets of nondiabetic Zucker fatty rats. *J. Biol. Chem.* **277**, 39163-39168.

Lowell, B. B. and Shulman, G. I. (2005). Mitochondrial dysfunction and type 2 diabetes. *Science* **307**, 384-387.

Maedler, K., Sergeev, P., Ris, F., Oberholzer, J., Joller-Jemelka, H. I., Spinas, G. A., Kaiser, N., Halban, P. A. and Donath, M. Y. (2002). Glucose-induced beta cell production of IL-1beta contributes to glucotoxicity in human pancreatic islets. *J. Clin. Invest.* **110**, 851-860.

Marchetti, P., Bugliani, M., Lupi, R., Marselli, L., Masini, M., Boggi, U., Filipponi, F., Weir, G. C., Eizirik, D. L. and Cnop, M. (2007). The endoplasmic reticulum in pancreatic beta cells of type 2 diabetes patients. *Diabetologia* **50**, 2486-2494.

Medina-Gomez, G., Gray, S. L., Yetukuri, L., Shimomura, K., Virtue, S., Campbell, M., Curtis, R. K., Jimenez-Linan, M., Blount, M., Yeo, G. S. et al. (2007). PPAR gamma 2 prevents lipotoxicity by controlling adipose tissue expandability and peripheral lipid metabolism. *PLoS Genet.* **3**, e64.

Melloul, D., Marshak, S. and Cerasi, E. (2002). Regulation of insulin gene transcription. *Diabetologia* **45**, 309-326.

Nakatani, Y., Kaneto, H., Kawamori, D., Yoshiuchi, K., Hatazaki, M., Matsuoka, T. A., Ozawa, K., Ogawa, S., Hori, M., Yamasaki, Y. et al. (2005). Involvement of endoplasmic reticulum stress in insulin resistance and diabetes. *J. Biol. Chem.* **280**, 847-851.

Olson, L. K., Redmon, J. B., Towle, H. C. and Robertson, R. P. (1993). Chronic exposure of HIT cells to high glucose concentrations paradoxically decreases insulin gene transcription and alters binding of insulin gene regulatory protein. *J. Clin. Invest.* **92**, 514-519.

Poitout, V. and Robertson, R. P. (2002). Minireview: Secondary beta-cell failure in type 2 diabetes-a convergence of glucotoxicity and lipotoxicity. *Endocrinology* **143**, 339-342.

Prentki, M. and Corkey, B. E. (1996). Are the beta-cell signaling molecules malonyl-CoA and cystolic long-chain acyl-CoA implicated in multiple tissue defects of obesity and NIDDM? *Diabetes* **45**, 273-283.

Prentki, M., Joly, E., El-Assaad, W. and Roduit, R. (2002). Malonyl-CoA signaling, lipid partitioning, and glucolipotoxicity: role in beta-cell adaptation and failure in the etiology of diabetes. *Diabetes* **51 Suppl. 3**, S405-S413.

Roduit, R., Nolan, C., Alarcon, C., Moore, P., Barbeau, A., Delghingaro-Augusto, V., Przybykowski, E., Morin, J., Masse, F., Massie, B. et al. (2004). A role for the malonyl-CoA/long-chain acyl-CoA pathway of lipid signaling in the regulation of insulin secretion in response to both fuel and nonfuel stimuli. *Diabetes* **53**, 1007-1019.

Saxena, N. K., Vertino, P. M., Anania, F. A. and Sharma, D. (2007). Leptin-induced growth stimulation of breast cancer cells involves recruitment of histone acetyltransferases and mediator complex to CYCLIN D1 promoter via activation of Stat3. *J. Biol. Chem.* **282**, 13316-13325.

Schulz, T. J., Zarse, K., Voigt, A., Urban, N., Birringer, M. and Ristow, M. (2007). Glucose restriction extends Caenorhabditis elegans life span by inducing mitochondrial respiration and increasing oxidative stress. *Cell Metab.* **6**, 280-293.

Shimano, H., Amemiya-Kudo, M., Takahashi, A., Kato, T., Ishikawa, M. and Yamada, N. (2007). Sterol regulatory element-binding protein-1c and pancreatic beta-cell dysfunction. *Diabetes Obes. Metab.* **9 Suppl. 2**, 133-139.

Stumvoll, M., Tataranni, P. A., Stefan, N., Vozarova, B. and Bogardus, C. (2003). Glucose allostasis. *Diabetes* **52**, 903-909.

Takahashi, A., Motomura, K., Kato, T., Yoshikawa, T., Nakagawa, Y., Yahagi, N., Sone, H., Suzuki, H., Toyoshima, H., Yamada, N. et al. (2005). Transgenic mice overexpressing nuclear SREBP-1c in pancreatic beta-cells. *Diabetes* **54**, 492-499.

Wang, H., Kouri, G. and Wollheim, C. B. (2005). ER stress and SREBP-1 activation are implicated in beta-cell glucolipotoxicity. *J. Cell Sci.* **118**, 3905-3915.

Wellen, K. E. and Hotamisligil, G. S. (2005). Inflammation, stress, and diabetes. *J. Clin. Invest.* **115**, 1111-1119.

Yamashita, T., Eto, K., Okazaki, Y., Yamashita, S., Yamauchi, T., Sekine, N., Nagai, R., Noda, M. and Kadowaki, T. (2004). Role of uncoupling protein-2 up-regulation and triglyceride accumulation in impaired glucose-stimulated insulin secretion in a beta-cell lipotoxicity model overexpressing sterol regulatory element-binding protein-1c. *Endocrinology* **145**, 3566-3577.

Yin, F., Wakino, S., Liu, Z., Kim, S., Hsueh, W. A., Collins, A. R., Van Herle, A. J. and Law, R. E. (2001). Troglitazone inhibits growth of MCF-7 breast carcinoma cells by targeting G1 cell cycle regulators. *Biochem. Biophys. Res. Commun.* **286**, 916-922.

PUBLICATION V

PPAR gamma 2 prevents lipotoxicity by controlling adipose tissue expandability and peripheral lipid metabolism

In: PloS Genetics 2007.
Vol. 3, Issue 4, e64.

PPAR gamma 2 Prevents Lipotoxicity by Controlling Adipose Tissue Expandability and Peripheral Lipid Metabolism

Gema Medina-Gomez[1], Sarah L. Gray[1], Laxman Yetukuri[2], Kenju Shimomura[3], Sam Virtue[1], Mark Campbell[1], R. Keira Curtis[1], Mercedes Jimenez-Linan[1], Margaret Blount[1], Giles S. H. Yeo[1], Miguel Lopez[1], Tuulikki Seppänen-Laakso[2], Frances M. Ashcroft[3], Matej Orešič[2], Antonio Vidal-Puig[1*]

1 Department of Clinical Biochemistry, Histopathology, University of Cambridge/Addenbrooke's Hospital, Cambridge, United Kingdom, 2 Technical Research Centre of Finland (VTT), Espoo, Finland, 3 University Laboratory of Physiology, University of Oxford, Oxford, United Kingdom

Peroxisome proliferator activated receptor gamma 2 (PPARg2) is the nutritionally regulated isoform of PPARg. Ablation of PPARg2 in the ob/ob background, PPARg2$^{-/-}$ Lep^{ob}/Lep^{ob} (POKO mouse), resulted in decreased fat mass, severe insulin resistance, β-cell failure, and dyslipidaemia. Our results indicate that the PPARg2 isoform plays an important role, mediating adipose tissue expansion in response to positive energy balance. Lipidomic analyses suggest that PPARg2 plays an important antilipotoxic role when induced ectopically in liver and muscle by facilitating deposition of fat as relatively harmless triacylglycerol species and thus preventing accumulation of reactive lipid species. Our data also indicate that PPARg2 may be required for the β-cell hypertrophic adaptive response to insulin resistance. In summary, the PPARg2 isoform prevents lipotoxicity by (a) promoting adipose tissue expansion, (b) increasing the lipid-buffering capacity of peripheral organs, and (c) facilitating the adaptive proliferative response of β-cells to insulin resistance.

Citation: Medina-Gomez G, Gray SL, Yetukuri L, Shimomura K, Virtue S, et al. (2007) PPAR gamma 2 prevents lipotoxicity by controlling adipose tissue expandability and peripheral lipid metabolism. PLoS Genet 3(4): e64. doi:10.1371/journal.pgen.0030064

Introduction

An adipocentric view of the Metabolic Syndrome (MS) considers obesity as the major factor leading to insulin resistance in peripheral metabolic tissues. However, the link between obesity and insulin resistance is complex, as indicated by the fact that some extremely obese people are glucose tolerant, while others with a mild degree of obesity develop severe insulin resistance and diabetes. This suggests that the absolute amount of fat stored may not be the most important factor determining the relationship between obesity and insulin resistance. Recent work showing the complexity of the molecular mechanisms controlling adipogenesis [1,2] suggests that adipose tissue expandability may be an important factor linking obesity, insulin resistance, and associated comorbidities.

There are two mechanisms that have been proposed to explain how expansion of the adipose tissue stores affects insulin sensitivity. One mechanism suggests that increased adiposity induces a chronic inflammatory state characterized by increased cytokine production by adipocytes and/or from macrophages infiltrating adipose tissue. Cytokines produced by these adipocytes or macrophages may directly antagonise insulin signalling [3,4]. A second nonexclusive hypothesis is lipotoxicity. The lipotoxic hypothesis states that if the amount of fuel entering a tissue exceeds its oxidative or storage capacity, toxic metabolites that inhibit insulin action are formed [5–8]. Of particular relevance to this article, lipid metabolites, such as ceramides and diacylglycerol (DAG) or reactive oxygen species generated from hyperactive oxidative pathways, have been shown to inhibit insulin signalling and to induce apoptosis [9–11].

The nuclear receptor peroxisome proliferator activated receptor gamma (PPARg) is critically required for adipogenesis and insulin sensitivity [12–15]. There are two PPARg isoforms, PPARg1 and PPARg2. PPARg1 is expressed in many tissues and cell types, including white and brown adipose tissue, skeletal muscle, liver, pancreatic β-cells, macrophages, colon, bone, and placenta [16]. Under physiological conditions, expression of PPARg2, the other splice variant, is restricted to white and brown adipose tissue [16,17]. In adipose tissue PPARg is the key regulator of adipogenesis. PPARg2 is the more adipogenic PPARg isoform in vitro, it is also the isoform regulated transcriptionally by nutrition [17–20]. Although under physiological conditions expression of PPARg2 is limited to adipose tissues, we have shown that PPARg2 is ectopically induced in liver and skeletal muscle in response to overnutrition or genetic obesity [2,18]. De novo expression of PPARg2 in liver and muscle in obesity suggests that PPARg2 may have a role in insulin resistance and

Editor: Gregory S. Barsh, Stanford University School of Medicine, United States of America

Received September 13, 2006; **Accepted** March 7, 2007; **Published** April 27, 2007

Abbreviations: DAG, diacylglycerol; GLUT, glucose transporter; H and E, haematoxylin and eosin; ITT, insulin tolerance test; LC/MS, liquid chromatography/mass spectrometry; MS, Metabolic Syndrome; PPARg, peroxisome proliferator activated receptor gamma; PPARg2, peroxisome proliferator activated receptor gamma 2; PPARGC1a, PPARg coactivator 1 alpha; *Scd1*, stearoyl-coenzyme A desaturase 1; TAG, triacylglycerol; WAT, white adipose tissue; WT, wild type

* To whom correspondence should be addressed. E-mail: ajv22@cam.ac.uk

Author Summary

It is known that obesity is linked to type 2 diabetes, however how obesity causes insulin resistance and diabetes is not well understood. Some extremely obese people are not diabetic, while other less obese people develop severe insulin resistance and diabetes. We believe diabetes occurs when adipose tissue becomes "full," and fat overflows into other organs such as liver, pancreas, and muscle, causing insulin resistance and diabetes. Peroxisome proliferator activated receptor gamma (PPARg) is essential for the development of adipose tissue and control of insulin sensitivity. PPARg2 is the isoform of PPARg regulated by nutrition. Here we investigate the role of PPARg2 under conditions of excess nutrients by removing the PPARg2 isoform in genetically obese mice, the POKO mouse. We report that removing PPARg2 decreases adipose tissue's capacity to expand and prevents the mouse from making as much fat as a normal obese mouse, despite eating similarly. Our studies suggest that PPARg plays an important antitoxic role when it is induced in liver, muscle, and beta cells by facilitating deposition of fat as relatively harmless lipids and thus prevents accumulation of toxic lipid species. We also show that PPARg2 may be involved in the adaptive response of beta cells to insulin resistance.

lipotoxicity in these tissues. Little in vivo research into the metabolic roles for the specific isoforms of PPARg has been carried out, with the studies so far focusing almost exclusively on adipose tissue [2,13,21,22]. PPARg (both isoforms) deletions have been generated in most major metabolic tissues. Liver-specific deletion of both PPARg isoforms caused an impairment in insulin sensitivity, particularly when challenged by different genetic backgrounds (lipoatrophic or leptin-deficiency) [23,24]. The effect of ablating both PPARg isoforms in muscle produced controversial results, with two groups reporting different effects on insulin sensitivity [25,26]. The role of PPARg in pancreatic β-cells is unclear, primarily due to its low expression under physiological conditions [27–29] and secondly because ablation of both PPARg isoforms in β-cells did not result in a metabolic phenotype. However PPARg may play a role in β-cell hyperplasia in response to insulin resistance, an idea supported by the fact that mice that lack PPARg in β-cells do not expand their β-cells mass in response to a high-fat diet [30]. More recently, it has been shown that heterozygous PPARg-deficient mice develop impaired insulin secretion, which is associated with increased islet triacylglycerol (TAG) content [31].

Here we investigate the physiological relevance of PPARg2 under conditions of positive energy balance by ablating PPARg2 in ob/ob mice. We use a new approach that integrates traditional physiological phenotyping with advanced lipidomic technology and transcriptomics. Our results indicate that in the context of positive energy balance, the absence of PPARg2 results in a major metabolic failure. Furthermore, we provide evidence that control of adipose tissue expansion by PPARg2 may be an important variable linking positive energy balance to its metabolic complications including insulin resistance, β-cell failure, and dyslipidaemia. Similarly, our lipidomic results indicate that induction of PPARg2 in nonadipose tissues should be considered as a physiological adaptation that prevents the toxic effects produced by excess nutrients. This antilipotoxic effect of PPARg2 is achieved by increasing the lipid-buffering capacity of peripheral organs and facilitating β-cell hyperplasia in response to insulin resistance.

Results

Ablation of PPARg2 in Ob/Ob Mice (POKO Mouse) Prevents Adipose Tissue Expansion in Response to Positive Energy Balance

PPARg2$^{-/-}$ Lep^{ob}/Lep^{ob} mice with genetic ablation of the PPARg2 isoform on the obese hyperphagic ob/ob background (POKO) were generated. Matings of PPARg2$^{+/-}$ Lep^{ob}/Lep^{+} mice followed the expected Mendelian distribution (Fisher's test = 0.074 and 0.135 for males and females, respectively). PPARg1 gene expression in white adipose tissue (WAT) from five-week-old POKO mice was similar to PPARg2 KO mice and was not significantly different from wild-type (WT) mice (Figure S1).

Figure 1A shows growth curves for male and female mice of four genotypes (WT, PPARg2 KO, ob/ob, and POKO mice) over a 12-week period. At birth, the body weight of male and female POKO mice was indistinguishable from other genotypes (unpublished data). The ob/ob mice quickly became heavier than their WT littermates, with significantly elevated body weight by four and six weeks of age in female and male mice, respectively. However, the POKO mice did not become obese, and their body weight remained close to WT and PPARg2 KO body weights mice during the 12-week study.

POKO mice were as hyperphagic (Figure 1B) as the ob/ob mice but drank far more water compared with ob/ob littermates (81.85 ± 15.14 versus 9.05 ± 2.32 ml/70 h, $p < 0.01$, female POKO versus ob/ob, $n = 4$ at 20 wk) (Figure S2A). Dual-energy X-ray absorptiometry analysis at 20 wk (Figure 1C) confirmed that female POKO mice had slightly increased fat content (4%) compared to WT and PPARg2 KO mice, but significantly reduced fat mass compared to the 40% increase observed in ob/ob mice. At the age of 20 wk, POKO and ob/ob mice had a trend to a decreased total locomotor activity during dark and light cycles compared with the WT and PPARg2 KO mice over the 72-h period. However POKO had similar total locomotor activity compared with ob/ob mice (Figure S2B).

At six weeks of age, female POKO mice consumed a similar amount of oxygen as ob/ob mice (vO_2 = 25.06 ± 0.89 versus 23.10 ± 0.99 ml/kg bodyweight $^{0.75}$/min, $p = 0.07$ POKO versus ob/ob, n = 6–8) showing a lower respiratory exchange ratio (0.916 ± 0.011 versus 0.952 ± 0.007, $p = 0.01$, female POKO versus ob/ob) in the fed state, but similar respiratory exchange ratio in the fasted state (0.73 ± 0.014 versus 0.75 ± 0.018, p-value = 0.59 POKO versus ob/ob mice). Water intake was already significantly increased in POKO compared to ob/ob mice (13.59 ± 1.88 versus 8.15 ± 0.89 ml/d, p-value < 0.05, POKO versus ob/ob). Furthermore, levels of glucose in urine were higher in POKO mice compared with ob/ob mice (403.4 ± 49.2 versus 34.13 ± 13.5 mMol/l, POKO versus ob/ob mice, p-value = 0.001), showing an energy loss of 15.43 ± 3.06 kJ/d through urine compared with 0.70 ± 0.19 kJ/d in ob/ob mice. At this age, POKO mice showed similar locomotor activity compared with the ob/ob mice during the day, but increased locomotor activity during the night (Figure S2C).

Histomorphometric analysis of adipose tissue from 16-wk-old male mice revealed that POKO mice had fewer small adipocytes than the ob/ob mice (Figure 1D and 1E). This

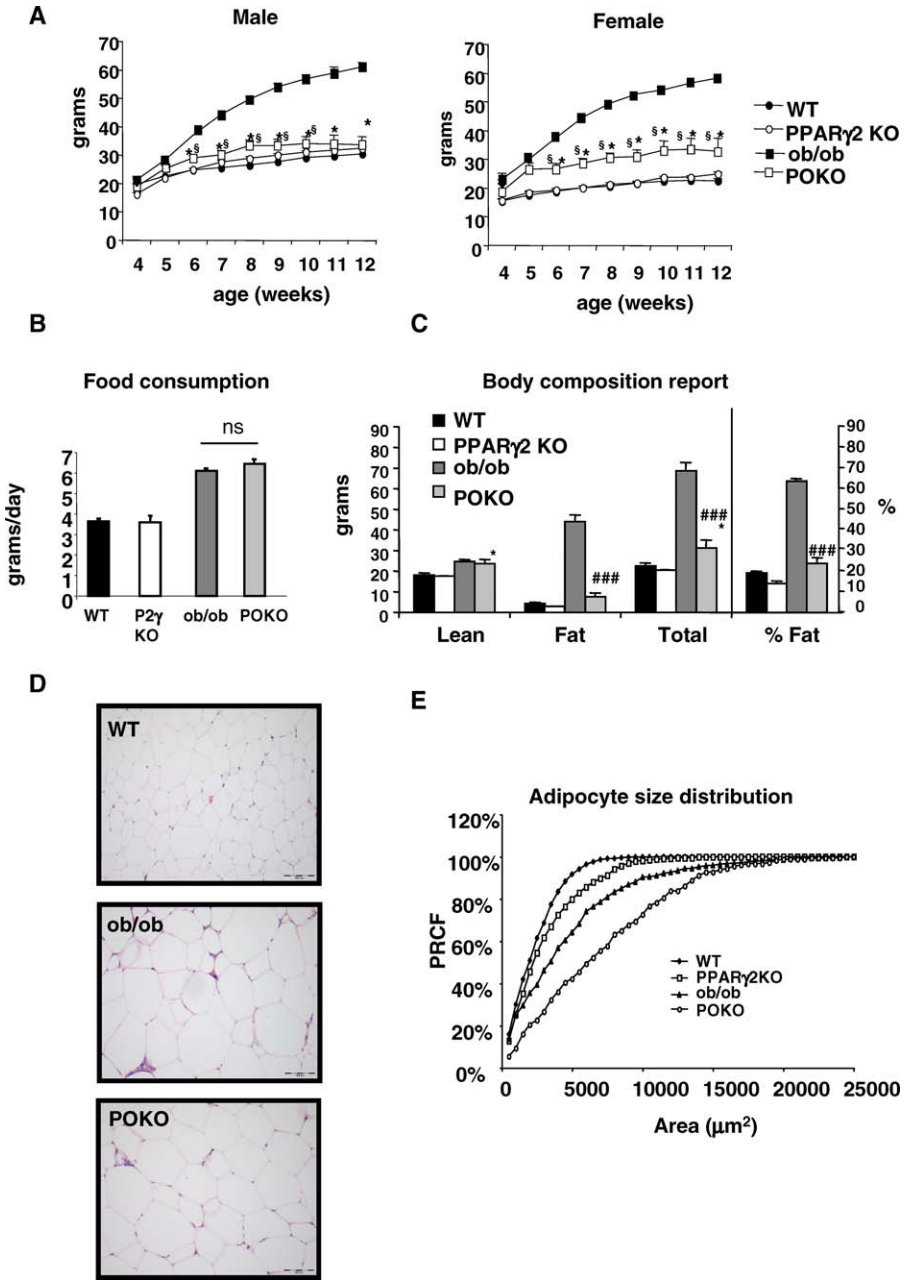


Figure 1. Physiological Characterisation of POKO Mouse

(A) Body weights (black circles, WT; black squares, ob/ob; white circles, PPARg2 KO; white squares, POKO) are shown for males (left) or females (right) ($n = 5$–12). *, $p < 0.05$ POKO versus ob/ob and §, $p < 0.01$ POKO versus WT.

(B) Food intake from 20-wk-old female mice ($n = 4$) is shown.

(C) Body composition analysis from 20-wk-old females is shown: WT, ob/ob, PPARg2 KO, and POKO mice fed chow diet mice ($n = 4$–7). *, $p < 0.05$ POKO versus WT and ###, $p < 0.001$ POKO versus ob/ob.

(D) Haematoxylin and eosin (H and E)-stained sections (10×) from epididymal WAT from 16-wk-old male WT, ob/ob, and POKO mice.

(E) Percent relative cumulative frequency analysis (PRCF) from epididymal WAT adipocytes from 16-wk-old male WT, ob/ob, PPARg2 KO, and POKO mice. ($n = 4$–5).

doi:10.1371/journal.pgen.0030064.g001

analysis of adipocyte size suggests that ablation of PPARg2 in the ob/ob background impairs the potential for adipocyte recruitment.

Early Insulin Resistance in POKO Mice Independent of Body Weight

As expected the reduced adipose tissue expandability of the POKO mouse was associated with severe insulin resistance. Surprisingly insulin resistance developed very early in life with elevated insulin levels and blood glucose compared to ob/ob mice (Table 1). We investigated whether peripheral insulin resistance and/or a severe defect in insulin secretion may cause hyperglycaemia in the POKO mouse. No differences in plasma glucose levels were detected three to five days after birth amongst the four genotypes for both genders (unpublished data). At weaning (three weeks of age) total

Table 1. Metabolic Parameters in Fed 4-Wk-Old Male and Female POKO, Ob/Ob, PPARg2 KO, and WT Mice

Mice	Parameters	Units	WT	PPARg2 KO	Ob/ob	POKO
Males 4 wk	Glucose	mMol/l	11.40 ± 0.56	10.65 ± 0.06	11.79 ± 0.66	27.80 ± 3.52*
	Insulin	μg/l	0.84 ± 0.11	1.06 ± 0.14	8.47 ± 2.04	11.26 ± 4.63
	FFA	mmol/l	0.63 ± 0.06	0.84 ± 0.05	0.43 ± 0.03	1.06 ± 0.09***
	TAGs	mmol/l	1.26 ± 0.27	1.5 ± 0.14	0.92 ± 0.08	3.33 ± 0.80**
	Adiponectin	μg/ml	19.17 ± 1.33	7.36 ± 1.57	25.34 ± 2.27	7.42 ± 1.16***
Females 4 wk	Glucose	mMol/l	9.07 ± 0.39	9.56 ± 1.04	10.50 ± 0.68	20.87 ± 1.79***
	Insulin	μg/l	0.83 ± 0.15	0.78 ± 0.12	16.98 ± 4.26	19.56 ± 2.98
	FFA	mmol/l	0.57 ± 0.06	0.76 ± 0.08	0.57 ± 0.08	0.76 ± 0.06
	TAGs	mmol/l	1.09 ± 0.09	1.49 ± 0.13	1.11 ± 0.13	4.51 ± 0.69***
	Adiponectin	μg/ml	28.41 ± 2.38	7.71 ± 0.07	38.12 ± 2.87	8.49 ± 2.10***

Values are mean ± standard error of mean. $n = 6/9$. *, $p < 0.05$; **, $p < 0.01$; and ***, $p < 0.001$ POKO versus ob/ob.
FFA, free fatty acid.
doi:10.1371/journal.pgen.0030064.t001

body weight was indistinguishable amongst the four genotypes, and blood glucose levels were similar in males and females (Figure 2A). However, by the age of four weeks, coincident with the change to a chow diet, male and female POKO mice developed severe hyperglycaemia compared to the other genotypes. Insulin plasma levels in the POKO mice at four weeks of age were increased compared to ob/ob mice (Table 1). Insulin resistance in POKO mice was confirmed by an insulin tolerance test (ITT) in four-week-old male and female mice (Figure 2B). Furthermore insulin resistance in adipose tissue was demonstrated by the extremely low levels of glucose transporter4 (GLUT4) protein in POKO adipose tissue when compared with GLUT4 levels in adipose tissue from ob/ob mice (Figure S3). Of note, insulin resistance in the POKO mice was associated with hypertriglyceridaemia as early as four-weeks of age (Table 1).

Adult POKO Mice are Hyperglycaemic and Have Low Plasma Insulin Levels

Given the early insulin resistance and hyperinsulinaemia in the young POKO mice, we expected to see increased insulin levels in mature POKO mice. At 16 weeks, male POKO mice exhibited severe hyperglycaemia in the fasted and fed states compared to littermate controls. Male POKO mice had inappropriately low levels of insulin (Table 2). A similar, but milder phenotype was also observed in POKO female mice (unpublished data). Of note, adult ob/ob mice compensated for their insulin resistance with increased insulin levels (Table 2). POKO mice also had hypertriglyceridaemia when compared to WT, ob/ob, or PPARg2 KO mice.

Impaired Beta-Cell Function in the POKO Mice

The inappropriately low insulin levels in the adult POKO mice suggested a defect in β-cells. Insulin resistance in ob/ob mice was compensated for by increasing pancreatic insulin secretion, islet number, and size (Figure 3A). However, despite being more insulin resistant than ob/ob mice, POKO mice did not increase their β-cell mass, resulting in lower plasma insulin levels than the ob/ob controls. Morphometric analysis of pancreatic sections from 16-week-old male mice confirmed that the islet-to-pancreas volume ratios were similar in the POKO, WT, and PPARg2 KO mice (0.023 ± 0.005, 0.013 ± 0.006, and 0.016 ± 0.005, respectively) and markedly increased in ob/ob mice (0.077 ± 0.017, $p < 0.01$ ob/ob versus POKO). Additionally, POKO mice had significantly decreased islet number and size (average area of islets POKO = 18.40 ± 2 mm^2) compared to ob/ob mice (ob/ob = 61.59 ± 8 mm^2). Insulin staining demonstrated that islets from POKO mice contained fewer insulin-positive cells than islets from ob/ob mice (Figure 3A). The normal cellular organization of the islet, abundant β-cells (insulin staining) in the centre of the islet and a rim of α-cells at the periphery (glucagon staining), was retained in the insulin resistant ob/ob mice but was disrupted in the islets of POKO mice (Figure 3A). Islets from POKO mice had decreased number of insulin positive β-cells when compared to islets from ob/ob mice and a scattered pattern of α-cells, which are morphological changes associated with islet remodelling in the context of β-cell failure. Gene expression analysis of islets from 16-week-old mice revealed decreased expression of pancreatic duodenal homeobox-1, insulin receptor substrate 2, *Glut2*, and insulin in islets from POKO mice when compared with those from WT or ob/ob (Figure S4).

The changes seen in the β-cells of POKO mice were not the result of an inherent failure of the β-cell to develop properly as indicated by histological studies of neonatal pancreas (day 3 to day 5) (unpublished data) and four-week-old pancreas (Figure 2C), showing no morphological differences in the size, number, or insulin staining of islets from POKO mice when compared to ob/ob controls.

Impaired Glucose-Stimulated Insulin Secretion in POKO Mouse Islets

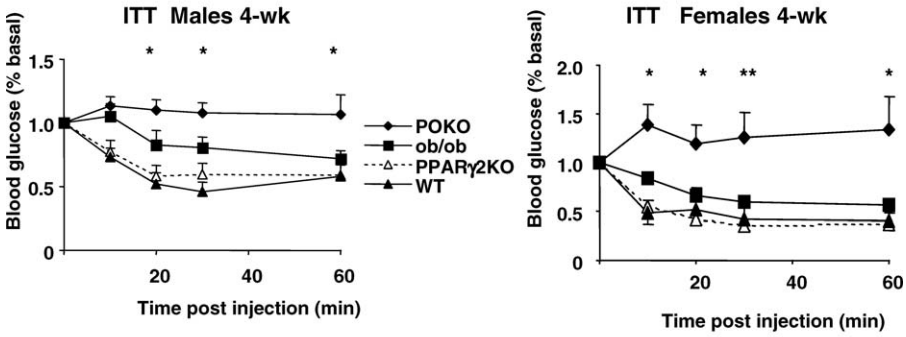
We measured glucose-stimulated insulin secretion in 16-week-old female POKO mice and their ob/ob littermates. Islets isolated from POKO mice were 30% smaller than those from ob/ob mice. Moreover, whereas normal islets were pure white with a smooth surface, islets from POKO mice were gray; their surface was irregular and required less time for collagenase digestion (only ten minutes instead of 30 minutes), suggesting that they were also more fragile.

Insulin content in islets from ob/ob mice was more than 30-fold greater than in those from POKO mice (Figure 3B). Insulin secretion from the islets of POKO mice was strikingly impaired compared to those of ob/ob mice, even when expressed relative to insulin content (Figure 3C). This was observed under basal (1 mM glucose) and stimulated (16 mM glucose, 16 mM glucose + tolbutamide) release.

A

	Females					
	Week 3		Week 4		Week 5	
	Weight (g)	Glucose (nmol/L)	Weight (g)	Glucose (nmol/L)	Weight (g)	Glucose (nmol/L)
WT	8.6±0.2	7.8±0.6	15.6±0.7	9.0±0.3	19.3±0.7	8.6±0.3
ob/ob	10.9±0.9	9.6±1.0	22.78±1.9	11.0±1.0	36.7±0.9	18.7±2.5
PPARγ2 KO	8.0±0.8	9.3±0.4	15.8±1.3	9.3±0.8	18.4±0.4	8.9±0.5
POKO	8.9±0.7	10.3±1.5	17.9±1.61	20.9±3.3**	27.8±1.0***	28.65±1.2*

B



C

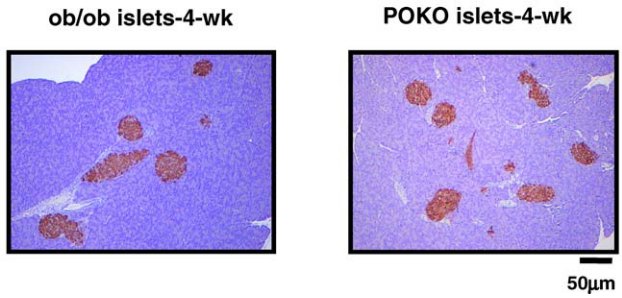


Figure 2. Early Insulin Resistance in POKO Mice Independent of Body Weight

(A) Body weight and plasma glucose levels from three, four, and five-week-old female WT, ob/ob, PPARg2 KO, and POKO. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$ POKO versus ob/ob.

(B) Plasma glucose levels during ITT on 4-wk-old male (left) and female (right) mice on chow diet (black triangle, WT; white triangle, PPARg2 KO; black square, ob/ob; black diamond, POKO) ($n = 7$). *, $p < 0.05$; **, $p < 0.01$ POKO versus ob/ob.

(C) Morphological analysis of H and E-stained sections (10×) in pancreas from 4-wk-old males ob/ob and POKO mice ($n = 5$).

doi:10.1371/journal.pgen.0030064.g002

Table 2. Metabolic Parameters in 16-wk-Old Male POKO, Ob/Ob, and WT Mice

Parameters	Units	WT	Ob/ob	POKO
Weight	g	36.0 ± 1.6	75.7 ± 4.5	40.4 ± 8.5*
Glucose fed	mMol/l	10.93 ± 1.45	15.27 ± 2.47	54.18 ± 6.5*,***
Glucose fasted	mMol/l	5.46 ± 0.52	10.74 ± 1.78	22.64 ± 3.98*,***
Insulin fed	μg/l	1.86 ± 0.49	46.63 ± 8.32	9.46 ± 1.84*,***
FFA	mmol/l	0.81 ± 0.10	0.88 ± 0.04	1.7 ± 0.43
Cholesterol	mmol/l	3.26 ± 0.05	6.40 ± 0.46	5.20 ± 0.49*,**
TAGs	mmol/l	1.68 ± 0.26	2.55 ± 0.68	9.06 ± 1.33*,***
Adiponectin	μg/ml	23.88 ± 1.08	13.74 ± 1.30	4.21 ± 0.98*,***

Values are mean ± standard error of mean. $n = 4/10$. *, $p < 0.001$ POKO versus ob/ob; **, $p < 0.05$ POKO versus WT; and ***, $p < 0.001$ POKO versus WT.

FFA, free fatty acid.

doi:10.1371/journal.pgen.0030064.t002

Decreased Steatosis in POKO Mice Compared to Ob/Ob Mice

As expected, the POKO mice had increased hepatic fat deposition compared to WT and PPARg2 KO mice (Table S1), but surprisingly the POKO mouse had much milder hepatosteatosis than the ob/ob mouse (Figure 3D), suggesting that ectopic expression of the PPARg2 isoform in the liver of ob/ob mice (see below), might contribute to the deposition of TAGs in the liver.

Ablation of PPARg2 Induces a Lipotoxic Lipid Profile in Adipose Tissue, Pancreatic Islets, Liver, and Skeletal Muscle

To investigate lipotoxicity as a potential pathogenic mechanism we used liquid chromatography/mass spectrometry (LC/MS) [32] to compare a broad spectrum of cellular lipids in the adipose tissue, pancreatic islets, liver, and

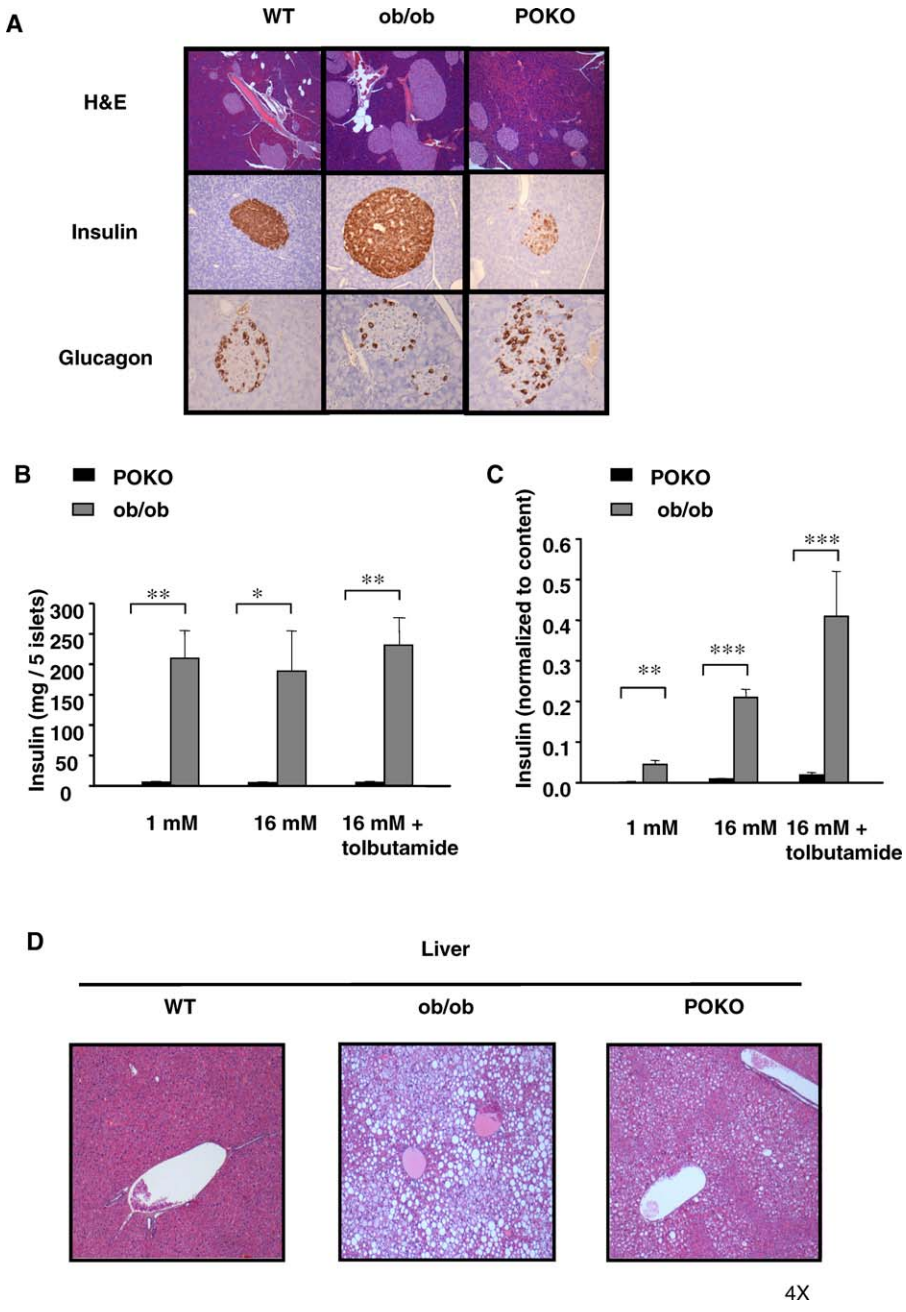


Figure 3. Impaired β-Cell Function and Hepatic Morphological Analyis in the POKO Mice

(A) H and E-stained sections (10×) and immunohistochemical (20×) analysis of insulin and glucagon in pancreas from 16-wk-old males WT, ob/ob, and POKO mice ($n = 5$).

(B) Insulin content of islets isolated from POKO (black bars), and ob/ob (grey bars) mice. Each data point is the mean of six samples each of five islets.

(C) Insulin secretion from islets isolated from POKO (black bars) and ob/ob (grey bars) mice in response to glucose (1, 16 mM) or glucose 16 mM + tolbutamide (200 μM). Data were collected from six samples each of five islets from three mice of each genotype. For each sample, insulin release was normalised to insulin content. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$ POKO versus ob/ob.

(D) H and E-stained sections (4×) in liver from 16-wk-old males WT, ob/ob, and POKO mice ($n = 5$).

doi:10.1371/journal.pgen.0030064.g003

skeletal muscle between the POKO mouse and controls (Protocol S1).

Adipose tissue from POKO mice has decreased TAG but increased DAG, ceramides, and other reactive lipid species associated with insulin resistance. Lipidomic analysis using LC/MS identified 74 molecular species differentially present in POKO, ob/ob, and WT mice (Protocol S1). POKO adipose tissue had decreased short chain TAGs compared to ob/ob adipose tissue (Protocol S1). Conversely, the concentration of DAGs was increased in the WAT of the POKO mice compared to ob/ob littermates. There was also an increased concentration of reactive lipid species in the WAT of POKO mice compared to that of ob/ob. The WAT of both POKO and ob/ob mice (Protocol S1) had increased levels of two ceramide species (with 16:0 and 24:1 fatty acid chains, respectively) and three proinflammatory lysophosphatidylcholine species [33]

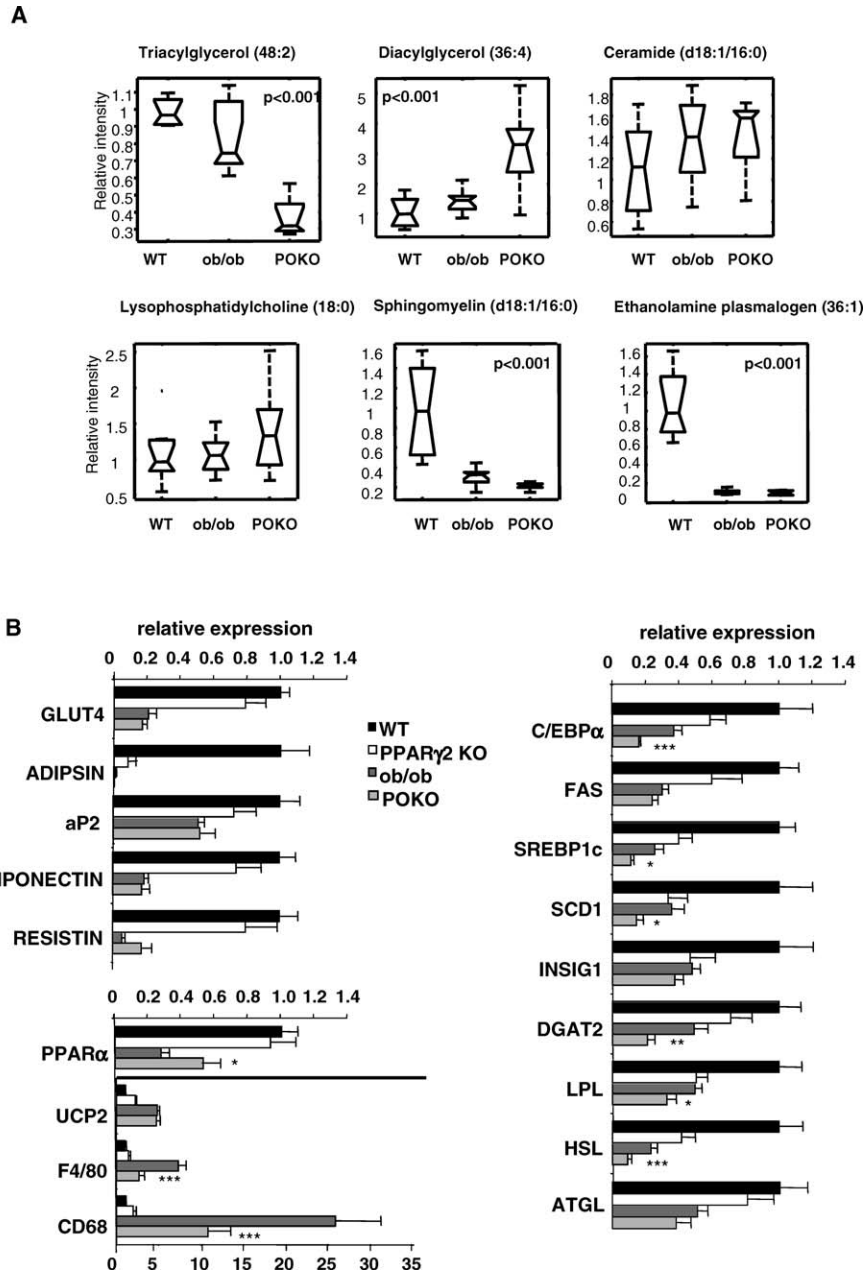


Figure 4. Lipidomic and Gene Expression Analysis of POKO WAT

(A) Lipidomic profiling of WAT from 16-wk-old males WT, ob/ob, and POKO mice.

(B) Adipose tissue mRNA levels from different genes from 16-wk-old male WT, PPARg2 KO, ob/ob, and POKO mice ($n = 6–8$). *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$ POKO versus ob/ob.

doi:10.1371/journal.pgen.0030064.g004

compared to WT mice. Partial least squares discriminant analysis indicated these changes in ceramides were greater in the POKO than ob/ob mice (Protocol S1). Sphingomyelin (d18:1/16:0), the precursor of ceramide (d18:1/16:0) and antioxidant ethanolamine plasmalogen (36:1) [34] were markedly decreased in POKO and ob/ob mice (Figure 4A).

Decreased TAG and accumulation of reactive lipid species in islets from POKO mice. Partial least-squares discriminant analysis of lipidomic profiles of isolated pancreatic islets of 16-week-old mice identified 44 lipid species accumulated at different concentrations in WT, PPARg2 KO, and POKO mice (Protocol S1). Short chain TAGs were decreased in islets from POKO and PPARg2 KO mice when compared to those from WT. This was associated with up-regulation of phosphatidylethanolamine (36:2), down-regulation of ethanolamine plasmalogen (36:2), and preferential accumulation of reactive lipid species, particularly of two ceramides (20:0 and 22:0 fatty acids) in islets from POKO mice (Figure 5A and Protocol S1).

Decreased TAG and increased reactive lipid species in liver of POKO mice. Multivariate analysis of lipidomic profiles (192 lipid species) revealed large changes between the POKO, PPARg2 KO, ob/ob, and WT genotypes (Protocol S1). These included decreased levels of short and medium chain TAGs

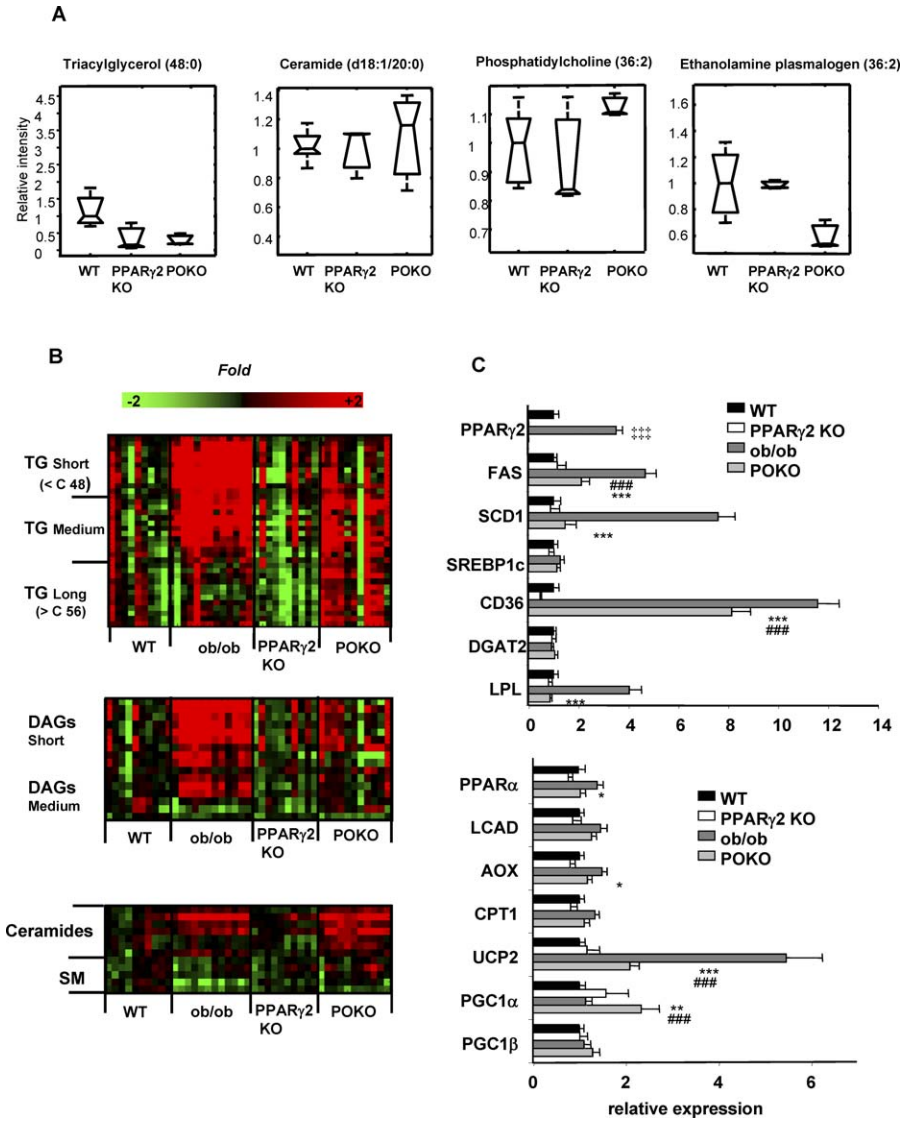


Figure 5. Lipidomic and Gene Expression Analysis in Islets and Liver from POKO Mice

Lipidomic profiling of islets (A) and liver (B) from 16-wk-old males WT, PPARg2 KO, ob/ob, and POKO mice. TG, TAGs; DAGs, diacylglycerols; SM, sphingomyelins. (C) Liver gene expression from 16-wk-old male WT, ob/ob, PPARg2 KO, and POKO mice fed chow ($n = 6$–8). *, $p < 0.05$; **, p, <0.01; ***, $p < 0.001$ POKO versus ob/ob; ###, $p < 0.001$ POKO versus WT; ‡‡‡, $p < 0.001$ ob/ob versus WT.

doi:10.1371/journal.pgen.0030064.g005

and DAGs (Figure 5B) in livers from POKO mice compared to those of ob/ob mice. Livers from POKO mice also had decrease levels of phosphatidylcholine lipid species (Protocol S1) utilised during the formation and secretion of very low density lipoproteins [35]. Conversely, POKO livers were enriched in ceramides compared to ob/ob livers, which correlated with the extent of increased levels of lysophosphatidylcholines in POKO and ob/ob mice (Protocol S1).

Decreased TAG and accumulation of reactive lipid species in POKO skeletal muscle. The same lipidomic pattern of decreased TAG and increased reactive lipid species previously observed in adipose tissue, β-cell, and liver was found to a milder degree in the skeletal muscle of POKO mice (Protocol S1). Briefly, when compared to ob/ob skeletal muscle, POKO skeletal muscle showed a decrease in very short-chain fatty acid TAGs and a slight decrease in levels of medium and long chain TAGs (Protocol S1). The skeletal muscle of POKO mice also had increased reactive lipids including ceramide (d18:1/18:0), DAGs, lysophosphatidylcholines, and sphingomyelins (precursors of ceramides) when compared to that of ob/ob mice.

Transcriptomic Analysis in POKO Mice Correlates with Lipidomic Changes

Given the lipotoxic profiles identified in the POKO mouse, we hypothesised changes in the expression of metabolic genes directly related to PPARg2 ablation and also compensatory changes in genes associated with cellular stress (Table S4).

Gene expression analysis in WAT. Target genes of PPARg such as *Glut4*, *adipsin*, *aP2*, and *adiponectin* were decreased to a larger extent in the WAT of five- and 16-week-old POKO mice than in PPARg2 KO mice (Figure S1 and Figure 4B). At five weeks of age, when differences in body fat between female WT, ob/ob, and POKO mice are only starting to

become evident, levels of GLUT4, aP2, and adiponectin mRNA levels were similar in WT and ob/ob mice, yet were markedly decreased in POKO mice. As the ob/ob mice aged (16 wk) and became obese and insulin resistant, the expression pattern of these PPARg targets in the WAT of ob/ob mice became similar to that of the POKO mice.

Results from the lipidomic analysis suggested major changes in the expression of genes involved in lipid metabolism (Figure 4B). Expression of stearoyl-coenzyme A desaturase 1 *(Scd1)* and sterol regulatory element-binding protein-1c (SREBP1c) were significantly lower in WAT from POKO mice compared to ob/ob mice. Furthermore, the decrease in TAGs and increased DAGs correlated with decreased expression of DAG acyltransferase 2, a key enzyme catalysing the final step in TAG synthesis, in the WAT of POKO mice compared with WAT from ob/ob mice. Again supporting the lipidomic profile, the expression of hormone-sensitive lipase, a rate-limiting enzyme for hydrolysis of diacylglycerides, was decreased in the WAT of POKO, PPARg2 KO, and ob/ob mice compared with WT mice, with the lowest levels observed in the POKO mice. Adipose triglyceride lipase levels were decreased in ob/ob and POKO compared with WT and PPARg2 KO mice, but without significant differences between ob/ob and POKO mice.

Oxidative stress has recently been suggested as a common mechanism of insulin resistance. Adipose tissue from POKO mice had increased oxidative stress compared to that of ob/ob mice as indicated by decreased gene expression levels of extracellular CuZn-superoxide dismutase, disruption of the glutathione pathway as indicated by decreased levels of gluthatione synthase, and increased levels of peroxidase and several gluthatione transferases (Table S2). We examined macrophage infiltration of adipose tissue as a potential marker of inflammation associated insulin resistance. Expression of CD68 and F4/80, both macrophage markers, was increased in the WAT of both POKO and ob/ob mice compared with WT and PPARg2 KO mice (Figure 4B). However their expression levels were lower in the POKO mice than the ob/ob mice suggesting that macrophage infiltration was not directly related to the exacerbated insulin resistance of the POKO mouse compared to the ob/ob mouse.

Gene expression in the POKO liver. Reduced hepatic steatosis accompanied by altered lipid profiles suggested that lack of hepatic ectopic expression of PPARg2 might be affecting lipid storage and metabolism in the liver of the POKO mice. Expression of genes involved in lipid metabolism in liver (Figure 5C) revealed that, proportional to the accumulation of TAGs in the liver, fatty acid synthase, *Scd1,* and the fatty acid translocase (FAT/CD36) were increased in ob/ob and POKO livers compared to WT mice and were significantly decreased in liver from POKO mice compared with liver from ob/ob mice. Other lipogenic PPARg target genes such as *Lpl* were also decreased in the POKO liver compared to the ob/ob mice. The ob/ob mice also had a compensatory increase in the expression of genes involved in β-oxidation (e.g., *Pparg, Lcad, Aox, Cpt1,* and *Ucp2*). Interestingly expression of these pro-oxidative genes was decreased in the liver of POKO mice when compared to that of ob/ob mice suggesting PPARg2 may contribute to their regulation [36].

Although β-cell failure could account for the severe hyperglycaemia observed in the POKO genotype, hepatic gluconeogenesis function might be affected. We observed a robust up-regulation of PPARg coactivator 1 alpha (PPARGC1a, also known as PGC1a) expression in the POKO liver compared with the WT and ob/ob mice. PPARGC1a is up-regulated in fasting and is thought to induce gluconeogenesis [37]. In parallel with the increase in PPARGC1a, microarray analysis revealed increased mRNA levels of the progluconeogenic genes phosphoenolpyruvate carboxykinase 1 *(Pepck1)* and glucose-6-phosphatase *(G6pc)* in the livers of POKO mice when compared to those of ob/ob mice (Table S2), suggesting hepatic gluconeogenesis may contribute to the hyperglycaemia observed in POKO mice.

Gene expression analysis in skeletal muscle of POKO mice. In 16-week-old POKO-mice skeletal muscle we observed down-regulation of *Srebp1c* and *Ppargc1a* and up-regulation of *Ucp2* expression in skeletal muscle from POKO mice compared to that of WT mice. Similarly, expression of *Lpl* and *Scd1* was down-regulated in the skeletal muscle of POKO mice when compared with that from ob/ob mice (Figure S5; Table S2). Gene set enrichment analysis of microarray data showed decreased expression of oxidative phosphorylation and mitochondrial components including electron transport chain complex components, in skeletal muscle from POKO mice when compared with that from ob/ob mice (Table S3).

Discussion

The link between obesity, insulin resistance, and diabetes while epidemiologically very clear is still not properly understood at a mechanistic level. An emerging concept is that the absolute amount of fat stored may be less important than the remaining storage capacity of the adipose tissue. Here we show that the PPARg2 isoform may be an important factor controlling obesity-induced comorbidities through two mechanisms: (a) by regulating nutritionally induced adipose tissue expandability and (b) when de novo expressed in nonadipose tissues, by allowing the storage of energy in the form of relatively harmless TAG species.

Previously we described the metabolic phenotype of the adult PPARg2 KO mouse [2], characterised by mild insulin resistance observed only in males. Given the greater adipogenic potency of PPARg2 compared with PPARg1 in vitro, we expected PPARg2 KO mice to have many more severe defects in adipose tissue than we observed, and therefore insulin sensitivity. As PPARg2 is the PPARg isoform regulated in response to nutrition and obesity [17–20], we hypothesised that PPARg2 would only become essential for adipose tissue function in the face of positive energy balance. The metabolic challenge we opted for was PPARg2 ablation in the obese (ob/ob) background (PPARg2$^{-/-}$ Lep^{ob}/Lep^{ob}, POKO mouse). The POKO mouse had severely decreased body-fat mass due to impaired adipose tissue expandability. Despite eating as much as an ob/ob mouse and expending a similar amount of energy, the POKO mouse was unable to store fat efficiently in its adipose tissue. This mismatch between increased energy availability and lack of adipose tissue expandability lead to a global metabolic failure characterised by severe insulin resistance, β-cell failure, and dyslipidaemia.

The observation of reduced fat mass and increased insulin resistance in the POKO mouse compared to the ob/ob mouse strongly supports two of our hypotheses. First, we hypoth-

esised that PPARg2 is required to recruit new adipocytes in overnutrition, but it is not required to make adipocytes during development. This is reflected by similar expression of aP2, a late marker of adipocyte differentiation, in POKO and ob/ob mice. The absence of small adipocytes was markedly different to other forms of lipodystrophy [38,39]. Additionally, and again in contrast with other lipodystrophic models that have markedly less adipose tissue than WT controls [38–40], the POKO mice had a percentage body fat that was similar (only 4% more) to WT and PPARg2 KO mice, as opposed to ob/ob mice, which had 40% fat as a proportion of body mass. This suggests that the remaining PPARg1 isoform is sufficient to support development of adipose tissue and fat deposition requirements of a lean mouse model. However, under conditions of positive energy balance, adipose tissue expandability mainly relies on the PPARg2 isoform. This idea is also suggested by the studies in heterozygous mice harbouring the murine equivalent of the human mutation *(P465L)* in PPARg on an ob/ob background [41]. These mice were able to accumulate fat and become obese even though showing a body mass 14% lower than ob/ob controls. In humans there is also evidence for a role for PPARg2. We have observed that metabolically healthy, nondiabetic, morbidly obese individuals have elevated levels of PPARg2 in their adipose tissue when compared to lean individuals [19]. Our second hypothesis, that the mismatch between energy availability and adipose tissue expandability is more important than fat mass itself as a predictor of insulin resistance, is also supported by our data. In fact the ob/ob mouse is much more obese than the POKO mouse but is much less insulin resistant. Furthermore, the POKO mice were already more insulin resistant than the ob/ob mice by the age of four weeks, with very low levels of GLUT4 in adipose tissue, before large differences in body weight developed, suggesting that the bioenergetic mismatch rather than the total amount of fat stored is important for the development of insulin resistance.

Although we hypothesised that the POKO mice would become insulin resistant, the degree of hyperglycaemia in these animals was in excess of what we expected. We found that the normal adaptive response of β-cells to insulin resistance did not occur in the POKO mice as indicated by the pathological changes observed by histology and the lack of β-cell hypertrophy. Although it has been shown that genetic background can affect the ability of ob/ob mice to undergo β-cell hypertrophy [42,43], we found that the ob/ob controls on our mixed 129Sv × C57BL/6J background underwent adaptive β-cell hyperplasia and hypertrophy, suggesting that the lack of PPARg2 was responsible for the failure of the POKO β-cells to adapt to insulin resistance. Interestingly the mass of pancreatic islets in POKO mice remained similar to the noninsulin resistant WT and PPARg2 KO mice. Furthermore, these defects in POKO β-cells did not appear to be the result of a developmental defect, as new born and four-week-old mice had morphologically normal islets.

The severe β-cell phenotype of the POKO mouse contrasts with the absence of hyperglycaemia observed in the pancreatic β-cell specific PPARg KO mouse [30]. However it should be kept in mind that in the β-cell specific PPARg KO mouse, the expression of PPARg and the lipid storage capacity of other tissues, most importantly adipose tissue, were not affected, and that insulin sensitivity was only mildly affected by high fat feeding in these mice when compared to the severe insulin resistance observed in POKO mice. Therefore the challenge to the pancreatic β-cells in this model was milder than in POKO mice. This is a clear example of how tissue-specific genetic manipulations are not always the best approach to understand the physiology of an organ in the context of the global energy homeostasis. The potential importance of the de novo expression of PPARg2 isoform in β-cells is also supported by the observation that humans harbouring the Pro12Ala mutation in PPARg2, a mutation that is located in the g2 isoform and makes PPARg2 less active, has only been associated with insulin deficiency and disease severity in obese individuals with type 2 diabetes [44].

The liver of the POKO mouse also displayed an unusual phenotype. We expected the POKO mice to have worse hepatosteatosis with increased triglyceride deposition in liver compared to ob/ob mice, because the POKO mice could not store fat in adipose tissue. However POKO mice had less hepatosteatosis than ob/ob mice suggesting that the PPARg2 isoform may directly contribute to facilitate triglyceride deposition in the liver.

A common mechanistic link for the phenotypes observed in the POKO liver and β-cell was not immediately obvious. To try to determine the role of PPARg2 in these locations we performed lipidomic and gene expression analyses of the adipose tissue, pancreatic islet, liver, and skeletal muscle of the POKO mouse. The lipid pattern of adipose tissue from POKO mice was characterised by decreased TAGs and increased DAGs in parallel with decreased gene expression of DGAT2, hormone-sensitive lipase, and adipose triglyceride lipase. This decrease in TAGs in the POKO adipose tissue was associated with increased levels of reactive lipid species and a gene expression profile suggestive of increased oxidative stress [45–49]. Although it has been described that oxidative stress and insulin resistance may be related to infiltration of adipose tissue by macrophages, resulting in a chronic state of inflammation [50–52], we did not observe increased macrophage infiltration in the adipose tissue of POKO mice compared to that of ob/ob mice.

Lipidomic analysis of POKO derived islets also showed decreased levels of triacyl and DAGs and increased levels of ceramides, suggesting that PPARg2 may contribute to increasing the lipid-buffering capacity of β-cells by promoting formation of TAGs and thus preventing lipotoxic insults. Liver and skeletal muscle lipidomics also showed reduced TAG and increased formation of reactive lipid species such as ceramides and lysophosphatidylcholines in POKO mice compared to ob/ob mice. This lipid profile was associated with impaired expression of pathways controlling de novo lipogenesis, transport of fatty acids, and beta oxidation in the POKO mice compared with the ob/ob mice. Of interest, *Ppargc1a* and other gluconeogenic genes were induced in the liver of POKO mice compared to that of ob/ob mice, suggesting a potential mechanism contributing to marked hyperglycaemia in POKO mice [53,54].

Overall, our lipidomic studies identify a remarkably similar pattern of changes in lipid species in the four tissues studied. The reduced adipose tissue mass and hepatosteatosis in the POKO mouse compared to the ob/ob mouse is explained by reduced levels of mature TAG in the POKO mouse. Similarly, ablation of PPARg2 resulted in accumulation of reactive lipid species implicated in causing insulin

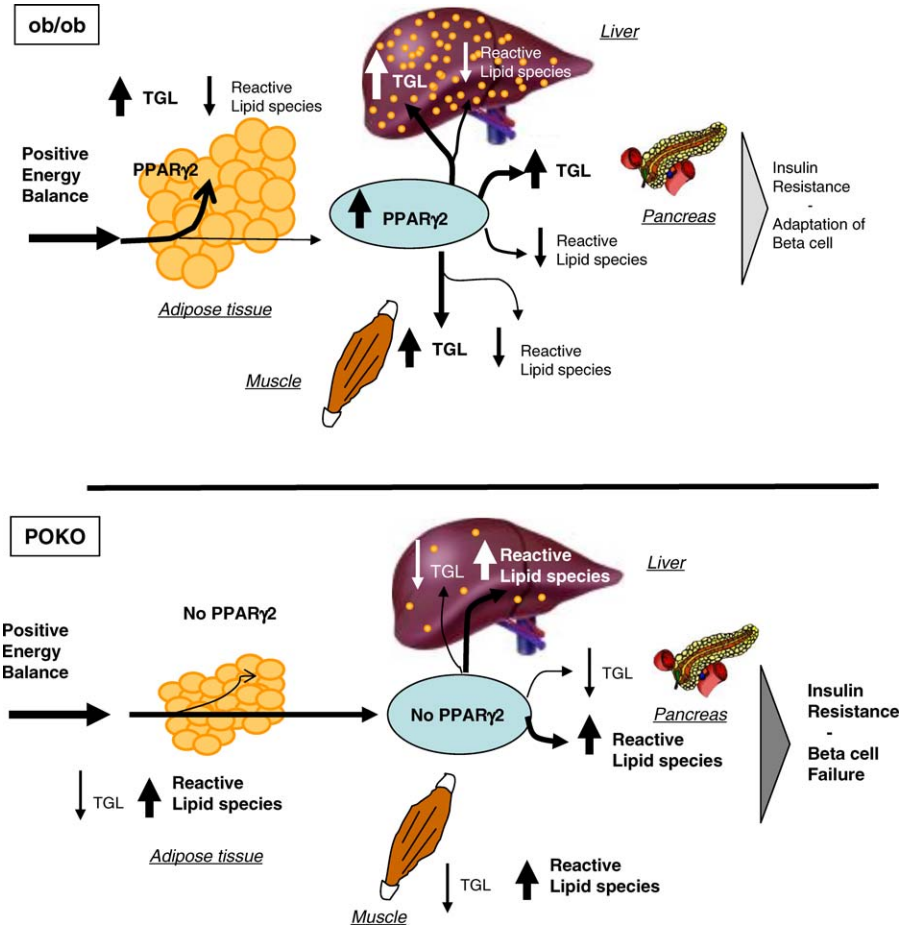


Figure 6. Storage of Lipids—Antilipotoxic Role of PPARg2

Antilipotoxic role of PPARg2 mediated by (a) expansion of adipose tissue and facilitation of triglyceride deposition and (b) facilitating deposition of fat in liver, skeletal muscle, and pancreas in the form of TAG. Ob/Ob mice can induce PPARg2 expression in liver, muscle, and β-cell, facilitating deposition of excess of energy in these organs in the form of TAG. Absence of inducibility of PPARg2 in POKO mouse liver, muscle, and β-cells results in increased deposition of reactive lipid species and decreased TAG, leading to marked insulin resistance and β-cell failure.

doi:10.1371/journal.pgen.0030064.g006

resistance, not only in adipose tissue, but also in other organs involved in whole-organism glucose metabolism. These results indicate that expression of PPARg2 in the pancreas, liver, and muscle of the ob/ob mouse may be performing a protective role, by increasing the capacity of these organs to buffer toxic lipid species by allowing accumulation of relatively harmless TAGs. The importance of this peripheral antilipotoxic role of PPARg2 becomes more evident if we consider that POKO and ob/ob mice are under the same degree of positive energy balance as determined by similar food intake, locomotor activity, and energy expenditure, that both models lack leptin, and that the only difference between ob/ob and POKO mice is the presence or absence of PPARg2. Given the decreased adipose tissue expandability of the POKO mice compared to ob/ob, it was anticipated that, as in the liver, muscle, or β-cells of lipodistrophic mice, the POKO mouse would accumulate more fat than the ob/ob. However, our results clearly indicate that mice lacking PPARg2, despite massive nutrient availability, are unable to deposit TAG in peripheral tissues and instead accumulate reactive lipid species in these organs. Therefore the pathologies of the liver and β-cell observed in the POKO mouse may be a result of a common lipotoxic insult facilitated by the absence of PPARg2 (Figure 6).

In summary, in this study we provide new insights into the physiological relevance of the PPARg2 isoform and identify adipose tissue expandability as an important determinant of metabolic complications. Ablation of PPARg2 decreases adipose tissue expandability, but its pathophysiological effects only become relevant in the context of a mismatch between energy availability and adipose tissue expansion. We show that PPARg2 also plays protective role when expressed de novo in peripheral organs by increasing their capacity to buffer toxic lipids. Ablation of PPARg2 under conditions of positive energy balance determined by absence of leptin produced early development of severe insulin resistance, β-cell failure, diabetes, and hyperlipidaemia. Extrapolation of this model to humans may suggest that normal to overweight individuals with positive energy balance and inappropriately severe manifestations of the MS may have a defect in PPARg2 and/or alternative mechanisms that control adipose tissue expandability.

Materials and Methods

Generation of mice homozygous for PPARg2 KO and leptin deficiency (ob/ob). Mice heterozygous for a disruption in exon B1 of PPARg2 on a 129Sv background ($PPARg2^{+/-}$) [2] were crossed with heterozygous ob/ob (Lep^{ob}/Lep^{+}) mice on a C57Bl/6 background to obtain mice heterozygous for both the PPARg2 ablation and the leptin point mutation ($PPARg2^{+/-}$ Lep^{ob}/Lep^{+}). These mice were crossed to obtain the four experimental genotypes: WT ($PPARg2^{+/+}$ Lep^{+}/Lep^{+}), PPARg2 KO ($PPARg2^{-/-}$ Lep^{+}/Lep^{+}), ob/ob ($PPARg2^{+/+}$ Lep^{ob}/Lep^{ob}), and POKO ($PPARg2^{-/-}$ Lep^{ob}/Lep^{ob}). Genotyping for deletion of PPARg2 and the point mutation in the ob gene was performed by PCR using standard protocols [2,55].

Animal care. Animals were housed at a density of four animals per cage in a temperature-controlled room (20–22 °C) with 12-h light/dark cycles. Food and water were available ad libitum unless noted. All animal protocols used in this study were approved by the UK Home Office and the University of Cambridge.

Blood and urine biochemistry, food intake, and body composition analysis. Mice of the four experimental genotypes were placed at weaning (three weeks of age) on a normal chow diet (10% of calories derived from fat; D12450B, Research Diets, http://www.researchdiets.com). Enzymatic assay kits were used for determination of plasma FFAs (Roche, http://www.roche.com) and TAGs (Sigma-Aldrich, http://www.sigmaaldrich.com). Elisa kits were used for measurements of leptin (R & D Systems, http://www.rndsystems.com), insulin (DRG Diagnostics International Limited, http://www.drg-international.com), and adiponectin (B-Bridge International, http://www.b-bridge.com) according to manufacturers' instructions. Dual-energy X-ray absorptiometry (DEXA, Lunar Corporation, http://www.lunarcorp.com) was used to measure body composition; glucose in blood and in urine and food intake were monitored in the four experimental genotypes as previously shown [2].

Oxygen consumption, water intake, and locomotor activity. Oxygen was measured using an eight-chamber open-circuit oxygen-monitoring system attached to and sampled from the chambers of a Comprehensive Laboratory Animal Monitoring System (CLAMS; Columbus Instruments, http://www.colinst.com). Water consumed was also measured using CLAMS. Mice were housed individually in specially built Plexiglass cages maintained at 22 °C under an alternating 12:12-h light-dark cycle (light period 08:00–20:00). Sample air was sequentially passed through oxygen (O_2) and carbon dioxide (CO_2) sensors (Columbus Instruments) for determination of O_2 and CO_2 content. Mice were acclimatized to monitoring cages for 72 h before data collection. Mice were weighed before each trial. Ambulatory activity of individually housed mice was evaluated using an eight-cage rack OPTO-M3 Sensor system (Columbus Instruments). Cumulative ambulatory activity counts were recorded every 5 min throughout the light and dark cycles.

Calculations of energy lost in urine. Energy lost in urine was calculated accordingly as previously shown before [56] using the following calculations:

Energy lost in urine kJ/day = (glucose in urine [mMol/l]/1,000) × molecular weight glucose × (water intake [ml/day]/1,000) × E $density_{carb}$; E $density_{carb}$ = energy density related to oxidations within the body for carbohydrates as glucose = 15.76 kJ/g.

RNA preparation and real-time quantitative RT-PCR. Total RNA was isolated from islets and tissues samples according to the manufacturer's instructions (RNAeasy kit, Qiagen, http://www.qiagen.com) and STAT60 (Tel-Test, http://www.isotexdiagnostics.com/tel-test.html). Real-time quantitative PCR was performed using a TaqMan 7900 (Applied Biosystems, http://www.appliedbiosystems.com) according to standard protocols.

Western blot analyses. The tissue samples (40 μg) were subjected to SDS-PAGE on 8% polyacrylamide gels. Proteins were then electrophoretically transferred to polyvinylidene difluoride filters. After transferring, the filters were blocked with 5% nonfat dry milk in TBS-Tween 20 followed by incubation with primary GLUT4 and extracellular signal-regulated kinase 1/2 (ERK1/2) antibodies (Promega, http://www.promega.com) overnight. The bands were quantified by scanning densitometry.

Light microscopy and immunohistochemcal analysis. Tissue samples for morphological and immunohistochemcal analysis were prepared according to published protocols [2]. Morphometric analyses of adipose tissue and pancreas sections were acquired using a digital camera and microscope (Olympus BX41, http://www.olympus.com), and cell areas were measured using AnalySIS software (Soft Imaging System, http://www.soft-imaging.net). For adipose tissue, two fields from each section were analysed to obtain the mean cell-area per animal ($n = 5$ per genotype). The Computer Assisted Stereology Toolbox (CAST) 2.0 system from Olympus was used to perform all measurements in the pancreas according to published protocols [57].

Isolation and culture of pancreatic islets. The pancreas was injected via the bile duct with cold Hank's solution containing 0.4% (w/v) liberase (Roche). The pancreas was removed, digested for 15–30 min, and islets collected by handpicking. Isolated islets were cultured overnight in h-cell medium (SBMI 06, hcell technology, http://www.hcell.com) at 37 °C in 5% CO_2 in air. Islets were used the day after isolation for insulin secretion studies or RNA extraction.

Insulin secretion studies. Insulin secretion from isolated islets (five islets/well) was measured during 1-hr static incubations in Krebs—Ringer Buffer containing either 1 mM glucose, 16.7 mM glucose, or 16.7 mM glucose plus 200 μM tolbutamide in DMSO. The supernatants were assayed for insulin. Insulin content was extracted using 95:5 ethanol/acetic acid. Insulin was measured using a Mouse Insulin ELISA kit (Mercodia, http://www.mercodia.com). Islets were isolated from three mice of each genotype for these experiments. Thus, the data are the mean of three separate experiments, in which data were collected for each test solution from six samples each of five islets. For each sample, insulin release was normalised to insulin content.

ITT. ITTs on four-week-old mice were performed as previously published [58].

Lipid profiling. For WAT and muscle, the tissue sample (50 mg) was homogenized with 0.15 M sodium chloride (300 μl), and the lipids were extracted with 2 ml of chloroform: methanol (2:1) and used for LC/MS as previously described [2].

For liver and islets, an aliquot (20 μl for liver or 10 μl for islets) of an internal standard mixture (11 reference compounds at concentration level 8–10 μg/ml), 50 μl of 0.15 M sodium chloride (for liver), and chloroform:methanol (2:1) (200 μl for liver or 90 μl for islets) was added to the tissue sample (20–30 mg). The sample was homogenized, vortexed (2 min for liver or 15 s for islets), let to stand (1 h for liver, 20 min for islets), and centrifuged at 10,000 RPM for 3 min. From the separated lower phase, an aliquot was mixed with 10 μl of a labelled standard mixture (three stable isotope-labelled reference compounds at concentration level 9–11 μg/ml), and 0.5–1.0 μl injection was used for LC/MS analysis.

Total lipid extracts were analysed on a Waters Q-Tof Premier mass spectrometer (http://www.waters.com) combined with an Acquity Ultra Performance LC (UPLC). The column, which was kept at 50 °C, was an Acquity UPLC BEH C18 10 × 50 mm with 1.7 μm particles. The binary solvent system (flow rate 0.200 ml/min) included A, water (1% 1 M NH_4Ac, 0.1% HCOOH), and B, LC/MS grade (Rathburn, http://www.rathburn.co.uk) acetonitrile/isopropanol (5:2, 1% 1 M NH_4Ac, 0.1% HCOOH). The gradient started from 65% A/35% B, reached 100% B in 6 min, and remained there for the next 7 min. The total run time per sample, including a 5 min re-equilibration step, was 18 min. The temperature of the sample organizer was set at 10 °C.

Mass spectrometry was carried out on Q-Tof Premier (Waters) run in ESI+ mode. The data were collected over the mass range of m/z 300–1,200 with scan duration of 0.2 s. The source temperature was set at 120 °C, and nitrogen was used as desolvation gas (800 l/h) at 250 °C. The voltages of the sampling cone and capillary were 39 V and 3.2 kV, respectively. Reserpine (50 μg/l) was used as the lock spray reference compound (5 μl/min; 10 s-scan frequency).

Data processing was performed using the MZmine software [59]. Identification was performed based on an internal reference database of lipid species, or alternatively utilizing the tandem mass spectrometry. The statistical analyses were performed using Matlab (Mathworks, http://www.mathworks.com) and the Matlab library PLS Toolbox (Eigenvector Research, http://www.eigenvector.com).

Tandem mass spectrometry was used for the identification of selected lipid species. MS/MS runs were performed by using ESI+ mode, collision energy ramp from 15–30 V, and mass range starting from m/z 150. The other conditions were as shown in the Protocol S1.

Statistics. Results were expressed as mean ± standard error of mean. Statistical analysis was performed using a two-tailed unpaired t-test between appropriate pairs of groups, and significance declared if p-values were less than 0.05.

Supporting Information

Figure S1. Adipose Tissue and Liver Gene Expression

Found at doi:10.1371/journal.pgen.0030064.sg001 (39 KB PPT).

Figure S2. Water Consumed and Locomotor Activity

Found at doi:10.1371/journal.pgen.0030064.sg002 (38 KB PPT).

Figure S3. GLUT4 protein expression in WAT

Found at doi:10.1371/journal.pgen.0030064.sg003 (65 KB PPT).

Figure S4. Gene Expression in Islets

Found at doi:10.1371/journal.pgen.0030064.sg004 (25 KB PPT).

Figure S5. Gene Expression in Muscle

Found at doi:10.1371/journal.pgen.0030064.sg005 (32 KB PPT).

Protocol S1. POKO Mouse Model Lipidomics Dataset

Found at doi:10.1371/journal.pgen.0030064.sd001 (208 KB PDF).

Table S1. Tissue Weights of 16-Wk-Old Male POKO, Ob/Ob, PPARg2 KO, and WT mice

Found at doi:10.1371/journal.pgen.0030064.st001 (29 KB PPT).

Table S2. Microarray Data

Found at doi:10.1371/journal.pgen.0030064.st002 (105 KB DOC).

Table S3. Pathway Analysis from Microarray Data

Found at doi:10.1371/journal.pgen.0030064.st003 (112 KB XLS).

Table S4. Accession Numbers

GenBank (http://www.ncbi.nlm.nih.gov/Genbank) accession numbers for the genes and gene products discussed in this paper.

Found at doi:10.1371/journal.pgen.0030064.st004 (58 KB DOC).

Acknowledgments

Animal care and husbandry provided by J. Carter, S. Shelton, H. Wetsby, H. Williams, A. Kant, J.P. Whiting, and G. Bevan. We thank D. Lam, M. Dale, and K. Burling for their technical assistance. We also thank J. Skepper and P.M. Coan for their help with morphometry analysis in pancreas and Peter Murgatroyd for his help with oxygen consumption experiments. We acknowledge Paradigm Therapeutics (http://www.paradigm-therapeutics.co.uk) for generating the PPARg2 KO mouse.

Author contributions. GMG, SLG, FMA, and AVP conceived and designed the experiments. GMG, SLG, LY, KS, SV, MB, ML, and MO performed the experiments. GMG, SLG, LY, KS, SV, MC, RKC, MJL, TSL, and MO analysed the data. GMG, LY, KS, SV, MC, RKC, MB, and GSHY contributed reagents/materials/analysis tools. GMG and AVP wrote the paper.

Funding. This work was supported by the European Union FP6 Hepatic and Adipose Tissue and Functions in the Metabolic Syndrome (Hepadip) integrated program (http://www.hepadip.org) (LSHM-CT-2005–018734); Diabetes UK; Medical Research Council; Wellcome Trust Integrative Physiology program; Academy of Finland (grant number 111338); and Marie Curie International Reintegration Grant from the European Community.

Competing interests. The authors have declared that no competing interests exist.

References

1. Rosen ED, Spiegelman BM (2000) Molecular regulation of adipogenesis. Annu Rev Cell Dev Biol 16: 145–171.
2. Medina-Gomez G, Virtue S, Lelliott C, Boiani R, Campbell M, et al. (2005) The link between nutritional status and insulin sensitivity is dependent on the adipocyte-specific peroxisome proliferator-activated receptor-gamma2 isoform. Diabetes 54: 1706–1716.
3. Dandona P, Aljada A, Bandyopadhyay A (2004) Inflammation: The link between insulin resistance, obesity and diabetes. Trends Immunol 25: 4–7.
4. Wellen KE, Hotamisligil GS (2005) Inflammation, stress, and diabetes. J Clin Invest 115: 1111–1119.
5. Kraegen EW, Cooney GJ, Ye JM, Thompson AL, Furler SM (2001) The role of lipids in the pathogenesis of muscle insulin resistance and beta cell failure in type II diabetes and obesity. Exp Clin Endocrinol Diabetes 109 Suppl 2: S189–S201.
6. Lelliott C, Vidal-Puig AJ (2004) Lipotoxicity, an imbalance between lipogenesis de novo and fatty acid oxidation. Int J Obes Relat Metab Disord 28 Suppl 4: S22–S28.
7. Unger RH (2003) Minireview: Weapons of lean body mass destruction: The role of ectopic lipids in the metabolic syndrome. Endocrinology 144: 5159–5165.
8. McGarry JD (2002) Banting lecture 2001: Dysregulation of fatty acid metabolism in the etiology of type 2 diabetes. Diabetes 51: 7–18.
9. Lee Y, Hirose H, Ohneda M, Johnson JH, McGarry JD, et al. (1994) Beta-cell lipotoxicity in the pathogenesis of non-insulin-dependent diabetes mellitus of obese rats: Impairment in adipocyte-beta-cell relationships. Proc Natl Acad Sci U S A 91: 10878–10882.
10. Lupi R, Dotta F, Marselli L, Del Guerra S, Masini M, et al. (2002) Prolonged exposure to free fatty acids has cytostatic and pro-apoptotic effects on human pancreatic islets: Evidence that beta-cell death is caspase mediated, partially dependent on ceramide pathway, and Bcl-2 regulated. Diabetes 51: 1437–1442.
11. Unger RH, Orci L (2002) Lipoapoptosis: Its mechanism and its diseases. Biochim Biophys Acta 1585: 202–212.
12. Rosen ED, Sarraf P, Troy AE, Bradwin G, Moore K, et al. (1999) PPAR gamma is required for the differentiation of adipose tissue in vivo and in vitro. Mol Cell 4: 611–617.
13. Koutnikova H, Cock TA, Watanabe M, Houten SM, Champy MF, et al. (2003) Compensation by the muscle limits the metabolic consequences of lipodystrophy in PPAR gamma hypomorphic mice. Proc Natl Acad Sci U S A 100: 14457–14462.
14. Barak Y, Nelson MC, Ong ES, Jones YZ, Ruiz-Lozano P, et al. (1999) PPAR gamma is required for placental, cardiac, and adipose tissue development. Mol Cell 4: 585–595.
15. Spiegelman BM (1998) PPAR-gamma: Adipogenic regulator and thiazolidinedione receptor. Diabetes 47: 507–514.
16. Escher P, Braissant O, Basu-Modak S, Michalik L, Wahli W, et al. (2001) Rat PPARs: Quantitative analysis in adult rat tissues and regulation in fasting and refeeding. Endocrinology 142: 4195–4202.
17. Werman A, Hollenberg A, Solanes G, Bjorbaek C, Vidal-Puig AJ, et al. (1997) Ligand-independent activation domain in the N terminus of peroxisome proliferator-activated receptor gamma (PPARgamma). Differential activity of PPARgamma1 and –2 isoforms and influence of insulin. J Biol Chem 272: 20230–20235.
18. Vidal-Puig A, Jimenez-Linan M, Lowell BB, Hamann A, Hu E, et al. (1996) Regulation of PPAR gamma gene expression by nutrition and obesity in rodents. J Clin Invest 97: 2553–2561.
19. Vidal-Puig AJ, Considine RV, Jimenez-Linan M, Werman A, Pories WJ, et al. (1997) Peroxisome proliferator-activated receptor gene expression in human tissues. Effects of obesity, weight loss, and regulation by insulin and glucocorticoids. J Clin Invest 99: 2416–2422.
20. Ren D, Collingwood TN, Rebar EJ, Wolffe AP, Camp HS (2002) PPARgamma knockdown by engineered transcription factors: Exogenous PPARgamma2 but not PPARgamma1 reactivates adipogenesis. Genes Dev 16: 27–32.
21. Zhang J, Fu M, Cui T, Xiong C, Xu K, et al. (2004) Selective disruption of PPAR{gamma}2 impairs the development of adipose tissue and insulin sensitivity. Proc Natl Acad Sci U S A 101: 10703–10708.
22. He W, Barak Y, Hevener A, Olson P, Liao D, et al. (2003) Adipose-specific peroxisome proliferator-activated receptor gamma knockout causes insulin resistance in fat and liver but not in muscle. Proc Natl Acad Sci U S A 100: 15712–15717.
23. Gavrilova O, Haluzik M, Matsusue K, Cutson JJ, Johnson L, et al. (2003) Liver peroxisome proliferator-activated receptor gamma contributes to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. J Biol Chem 278: 34268–34276.
24. Matsusue K, Haluzik M, Lambert G, Yim SH, Gavrilova O, et al. (2003) Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty liver but aggravates diabetic phenotypes. J Clin Invest 111: 737–747.
25. Norris AW, Chen L, Fisher SJ, Szanto I, Ristow M, et al. (2003) Muscle-specific PPARgamma-deficient mice develop increased adiposity and insulin resistance but respond to thiazolidinediones. J Clin Invest 112: 608–618.
26. Hevener AL, He W, Barak Y, Le J, Bandyopadhyay G, et al. (2003) Muscle-specific Pparg deletion causes insulin resistance. Nat Med 9: 1491–1497.
27. Braissant O, Foufelle F, Scotto C, Dauca M, Wahli W (1996) Differential expression of peroxisome proliferator-activated receptors (PPARs): Tissue distribution of PPAR-alpha, -beta, and -gamma in the adult rat. Endocrinology 137: 354–366.
28. Dubois M, Pattou F, Kerr-Conte J, Gmyr V, Vandewalle B, et al. (2000) Expression of peroxisome proliferator-activated receptor gamma (PPAR-gamma) in normal human pancreatic islet cells. Diabetologia 43: 1165–1169.
29. Patane G, Anello M, Piro S, Vigneri R, Purrello F, et al. (2002) Role of ATP production and uncoupling protein-2 in the insulin secretory defect induced by chronic exposure to high glucose or free fatty acids and effects of peroxisome proliferator-activated receptor-gamma inhibition. Diabetes 51: 2749–2756.
30. Rosen ED, Kulkarni RN, Sarraf P, Ozcan U, Okada T, et al. (2003) Targeted elimination of peroxisome proliferator-activated receptor gamma in beta cells leads to abnormalities in islet mass without compromising glucose homeostasis. Mol Cell Biol 23: 7222–7229.
31. Matsui J, Terauchi Y, Kubota N, Takamoto I, Eto K, et al. (2004) Pioglitazone reduces islet triglyceride content and restores impaired glucose-stimulated insulin secretion in heterozygous peroxisome prolifer-

ator-activated receptor-gamma-deficient mice on a high-fat diet. Diabetes 53: 2844–2854.
32. Oresic M, Vidal-Puig A, Hanninen V (2006) Metabolomic approaches to phenotype characterization and applications to complex diseases. Expert Rev Mol Diagn 6: 575–585.
33. Takabe W, Kanai Y, Chairoungdua A, Shibata N, Toi S, et al. (2004) Lysophosphatidylcholine enhances cytokine production of endothelial cells via induction of L-type amino acid transporter 1 and cell surface antigen 4F2. Arterioscler Thromb Vasc Biol 24: 1640–1645.
34. Zoeller RA, Lake AC, Nagan N, Gaposchkin DP, Legner MA, et al. (1999) Plasmalogens as endogenous antioxidants: Somatic cell mutants reveal the importance of the vinyl ether. Biochem J 338 (Pt 3): 769–776.
35. Yao ZM, Vance DE (1988) The active synthesis of phosphatidylcholine is required for very low density lipoprotein secretion from rat hepatocytes. J Biol Chem 263: 2998–3004.
36. Patsouris D, Reddy JK, Muller M, Kersten S (2006) Peroxisome proliferator-activated receptor alpha mediates the effects of high-fat diet on hepatic gene expression. Endocrinology 147: 1508–1516.
37. Puigserver P, Rhee J, Donovan J, Walkey CJ, Yoon JC, et al. (2003) Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1alpha interaction. Nature 423: 550–555.
38. Shimomura I, Hammer RE, Richardson JA, Ikemoto S, Bashmakov Y, et al. (1998) Insulin resistance and diabetes mellitus in transgenic mice expressing nuclear SREBP-1c in adipose tissue: Model for congenital generalized lipodystrophy. Genes Dev 12: 3182–3194.
39. Ross SR, Graves RA, Spiegelman BM (1993) Targeted expression of a toxin gene to adipose tissue: Transgenic mice resistant to obesity. Genes Dev 7: 1318–1324.
40. Moitra J, Mason MM, Olive M, Krylov D, Gavrilova O, et al. (1998) Life without white fat: A transgenic mouse. Genes Dev 12: 3168–3181.
41. Gray SL, Nora ED, Grosse J, Manieri M, Stoeger T, et al. (2006) Leptin deficiency unmasks the deleterious effects of impaired peroxisome proliferator-activated receptor gamma function (P465L PPARgamma) in mice. Diabetes 55: 2669–2677.
42. Hummel KP, Dickie MM, Coleman DL (1966) Diabetes, a new mutation in the mouse. Science 153: 1127–1128.
43. Leiter EH (1981) The influence of genetic background on the expression of mutations at the diabetes locus in the mouse IV. Male lethal syndrome in CBA/Lt mice. Diabetes 30: 1035–1044.
44. Mori H, Ikegami H, Kawaguchi Y, Seino S, Yokoi N, et al. (2001) The Pro12 –>Ala substitution in PPAR-gamma is associated with resistance to development of diabetes in the general population: Possible involvement in impairment of insulin secretion in individuals with type 2 diabetes. Diabetes 50: 891–894.
45. Evans JL, Goldfine ID, Maddux BA, Grodsky GM (2002) Oxidative stress and stress-activated signaling pathways: A unifying hypothesis of type 2 diabetes. Endocr Rev 23: 599–622.
46. Haber CA, Lam TK, Yu Z, Gupta N, Goh T, et al. (2003) N-acetylcysteine and taurine prevent hyperglycemia-induced insulin resistance in vivo: Possible role of oxidative stress. Am J Physiol Endocrinol Metab 285: E744–E753.
47. Hildebrandt W, Hamann A, Krakowski-Roosen H, Kinscherf R, Dugi K, et al. (2004) Effect of thiol antioxidant on body fat and insulin reactivity. J Mol Med 82: 336–344.
48. Fridlyand LE, Philipson LH (2006) Reactive species and early manifestation of insulin resistance in type 2 diabetes. Diabetes Obes Metab 8: 136–145.
49. Bloch-Damti A, Bashan N (2005) Proposed mechanisms for the induction of insulin resistance by oxidative stress. Antioxid Redox Signal 7: 1553–1567.
50. Weisberg SP, McCann D, Desai M, Rosenbaum M, Leibel RL, et al. (2003) Obesity is associated with macrophage accumulation in adipose tissue. J Clin Invest 112: 1796–1808.
51. Xu H, Barnes GT, Yang Q, Tan G, Yang D, et al. (2003) Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. J Clin Invest 112: 1821–1830.
52. Qi C, Pekala PH (2000) Tumor necrosis factor-alpha-induced insulin resistance in adipocytes. Proc Soc Exp Biol Med 223: 128–135.
53. Herzig S, Long F, Jhala US, Hedrick S, Quinn R, et al. (2001) CREB regulates hepatic gluconeogenesis through the coactivator PGC-1. Nature 413: 179–183.
54. Yoon JC, Puigserver P, Chen G, Donovan J, Wu Z, et al. (2001) Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. Nature 413: 131–138.
55. Hirasawa T, Ohara T, Makino S (1997) Genetic typing of the mouse ob mutation by PCR and restriction enzyme analysis. Exp Anim 46: 75–78.
56. Elia M, Livesey G (1992) Energy expenditure and fuel selection in biological systems: The theory and practice of calculations based on indirect calorimetry and tracer methods. World Rev Nutr Diet 70: 68–131.
57. Gundersen HJ, Osterby R (1981) Optimizing sampling efficiency of stereological studies in biology: Or 'do more less well!' J Microsc 121: 65–73.
58. Vidal-Puig AJ, Grujic D, Zhang CY, Hagen T, Boss O, et al. (2000) Energy metabolism in uncoupling protein 3 gene knockout mice. J Biol Chem 275: 16258–16266.
59. Katajamaa M, Oresic M (2005) Processing methods for differential analysis of LC/MS profile data. BMC Bioinformatics 6: 179.

Series title, number and
report code of publication

VTT Publications 741
VTT-PUBS-741

Author(s)
Laxmana Rao Yetukuri

Title

Bioinformatics approaches for the analysis of lipidomics data

Abstract

The potential impact of lipid research has been increasingly realised both in disease treatment and prevention. Recent advances in soft ionization mass spectrometry (MS) such as electrospray ionization (ESI) have permitted parallel monitoring of several hundreds of lipids in a single experiment and thus facilitated lipidomics level studies. These advances, however, pose a greater challenge for bioinformaticians to handle massive amounts of information-rich MS data from modern analytical instruments in order to understand complex functions of lipids**.** The main aims of this thesis were to 1) develop bioinformatics approaches for lipid identification based on ultra performance liquid chromatography coupled to mass spectrometry (UPLC/MS) data, 2) predict the functional annotations for unidentified lipids, 3) understand the *omics* data in the context of pathways and 4) apply existing chemometric methods for exploratory data analysis as well as biomarker discovery.

A bioinformatics strategy for the construction of lipid database for major classes of lipids is presented using simplified molecular input line entry system (SMILES) approach. The database was annotated with relevant information such as lipid names including short names, SMILES information, scores, molecular weight, monoisotopic mass, and isotope distribution. The database was tailored for UPLC/MS experiments by incorporating the information such as retention time range, adduct information and main fragments to screen for the potential lipids. This database information facilitated building experimental tandem mass spectrometry libraries for different biological tissues.

Non-targeted metabolomics screening is often get plagued by the presence of unknown peaks and thus present an additional challenge for data interpretation. Multiple supervised classification methods were employed and compared for the functional prediction of class labels for unidentified lipids to facilitate exploratory analysis further as well as ease the identification process. As lipidomics goes beyond complete characterization of lipids, new strategies were developed to understand lipids in the context of pathways and thereby providing insights for the phenotype characterization. Chemometric methods such as principal component analysis (PCA) and partial least squares and discriminant analysis (PLS/DA) were utilised for exploratory analysis as well as biomarker discovery in the context of different disease phenotypes.

ISBN
978-951-38-7402-5 (soft back ed.)
978-951-38-7403-2 (URL: http://www.vtt.fi/publications/index.jsp)

Series title and ISSN	Project number
VTT Publications 1235-0621 (soft back ed.) 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)	

Date	Language	Pages
May 2010	English	75 p. + app. 106 p.

Name of project	Commissioned by

Keywords
Lipids, Lipidomics, Bioinformatics, Lipid pathways, High density lipoproteins, k-nearest neighbours, Liquid chromatography/mass spectrometry, Principal component analysis, Partial least squares and discriminant analysis, Obesity, Support vector machines, LipidDB

Publisher
VTT Technical Research Centre of Finland
P.O. Box 1000, FI-02044 VTT, Finland
Phone internat. +358 20 722 4520
Fax +358 20 722 4374

Lipid research has recently gained increased attention due to their involvement in numerous diseases including diabetes, obesity, atherosclerosis and Alzheimer's disease. Modern technological advancements in mass spectrometry allowed us to study several hundreds of lipids at a time. This level of investigation not only provides insights into the specific roles of lipid molecular species in the health and disease, but also facilitates in identifying potential biomarkers for prevention and treatment of human health. These developments, however, comes with a set of informatics challenges in terms of handling the data. This thesis mainly deals with some of the challenges associated with this kind of lipid research in the context of liquid chromatography/mass spectrometry methods. The presented informatics methods herein assist in identification of molecular species, their functional class prediction, and data interpretation in biological pathway context and data analysis

ISBN 978-951-38-7402-5 (soft back ed.)
ISSN 1235-0621 (soft back ed.)

ISBN 978-951-38-7403-2 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)