PAPER

Glycomics meets lipidomics—associations of *N*-glycans with classical lipids, glycerophospholipids, and sphingolipids in three European populations†

Wilmar Igl,‡[a] Ozren Polašek,‡[b] Olga Gornik,[c] Ana Knežević,[d] Maja Pučić,[d] Mislav Novokmet,[d] Jennifer Huffman,[e] Carsten Gnewuch,[f] Gerhard Liebisch,[f] Pauline M. Rudd,[g] Harry Campbell,[h] James F. Wilson,[h] Igor Rudan,[ijk] Ulf Gyllensten,[a] Gerd Schmitz[f] and Gordan Lauc[cd]

Recently, high-throughput technologies have been made available which allow the measurement of a broad spectrum of glycomics and lipidomics parameters in many samples. The aim of this study was to apply these methods and investigate associations between 46 glycan and 183 lipid traits measured in blood of 2041 Europeans from three different local populations (Croatia - VIS cohort; Sweden - NSPHS cohort; Great Britain - ORCADES cohort). *N*-glycans have been analyzed with High Performance Liquid Chromatography (HPLC) and lipids with Electrospray Ionization Tandem Mass Spectrometry (ESI-MS/MS) covering sterol lipids, glycerolipids, glycerophospholipids and sphingolipids in eight subclasses. Overall, 8418 associations were calculated using linear mixed effect models adjusted for pedigree, sex, age and multiple testing. We found 330 significant correlations in VIS. Pearson's correlation coefficient r ranged from -0.27 to 0.34 with corresponding p-values between 1.45×10^{-19} and 4.83×10^{-6}, indicating statistical significance. A total of 71 correlations in VIS could be replicated in NSPHS ($r = [-0.19; 0.35]$, $p = [4.16 \times 10^{-18}; 9.38 \times 10^{-5}]$) and 31 correlations in VIS were also found in ORCADES ($r = [-0.20; 0.24]$, $p = [2.69 \times 10^{-10}; 7.55 \times 10^{-5}]$). However, in total only 10 correlations between a subset of triantennary glycans and unsaturated phosphatidylcholine, saturated ceramide, and sphingomyelin lipids in VIS ($r = [0.18; 0.34]$, $p = [2.98 \times 10^{-21}; 1.69 \times 10^{-06}]$) could be replicated in both NSPHS and ORCADES. In summary, the results show strong and consistent associations between certain glycans and lipids in all populations, but also population-specific correlations which may be caused by environmental and genetic differences. These associations point towards potential interactive metabolic pathways.

Introduction

The recent development of high-throughput methods has enabled a new type of large-scale study which encompasses thousands of samples and hundreds of traits, such as nucleic acids, proteins, lipids and sugars.[1] While the progress in the fields of genomics and proteomics was rather fast, the development of lipidomics and glycomics substantially lagged behind. Reasons were the high diversity and complexity of lipids and glycans and the resulting difficulties in their quantification.

[a] *Department of Genetics and Pathology, Rudbeck Laboratory, University of Uppsala, SE-75185 Uppsala, Sweden*
[b] *Department of Public Health, Medical School, University of Split, Šoltanska 2, HR-21000 Split, Croatia*
[c] *University of Zagreb, Faculty of Pharmacy and Biochemistry, Ante Kovačića 1, 10000 Zagreb, Croatia. E-mail: glauc@pharma.hr; Fax: +385 1 639 4400*
[d] *Genos Ltd, Glycobiology Division, Planinska 1, HR-10000 Zagreb, Croatia*
[e] *MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, Scotland, Great Britain*
[f] *Institute for Clinical Chemistry and Laboratory Medicine, Regensburg University Medical Center, DE-93053 Regensburg, Germany*
[g] *Dublin-Oxford Glycobiology Lab., Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland*
[h] *Centre for Population Health Sciences and Institute of Genetics and Molecular Medicine, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, Scotland, Great Britain*
[i] *Department for Population Health Sciences, The University of Edinburgh Medical School, Teviot Place, Edinburgh, Great Britain*
[j] *Croatian Centre for Global Health, University of Split Medical School, Split, Croatia*
[k] *Gen-info Ltd, Ružmarinka 17, HR-10000 Zagreb, Croatia*
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c0mb00095g
‡ These authors contributed equally to the manuscript.

However, recently high-throughput methods have also been developed in the fields of lipidomics[2] and glycomics[3] which now enable exciting new insights into these biological traits.

At present, software tools that could explain the observed relations within complex data sets and model pathways and their interactions from a systems biology perspective[4,5] are still under development.[6,7] However, to get a better understanding of the functional relevance, studies examining simple associations between a broad range of omics traits with selected traits of medical importance have been conducted. For example, genome-wide association studies correlate common variants across the whole genome with traits of medical relevance, e.g. classical blood lipids.[8] These studies have contributed enormously to the identification of genes and to the understanding of their biological and medical relevance. However, omics studies mostly concentrate on a single omics discipline and selected traits of medical importance. Few studies exist which try to understand the interactions between two comprehensive sets of omics traits.[9] Therefore, we conducted the first glycome- and lipidome-wide association study to reveal interactions between these two major biological substance classes.

The glycome is defined as the entire set of glycans of an organism. Glycans are complex molecules which typically contain between 10 and 15 monosaccharides that are linked in a complicated manner. Glycosylation is the only post-translational modification that can produce substantial structural changes to proteins by attaching between two and five covalently bound glycans to proteins.[10] It is not template driven and the complexity of the glycoproteome is estimated to be several orders of magnitude greater than for the proteome itself.[11] Glycans play an important role in many metabolic processes, including protein folding, degradation and secretion, cell signaling, immune function and transcription.[12] Glycosylation is essential for multicellular life and its complete absence is embryonically lethal.[13] Dysregulation of glycosylation is implicated in a wide range of diseases, including cancer, diabetes, cardiovascular, congenital, immunological and infectious disorders.[14,15]

The lipidome is the entire set of lipids of an organism. Lipids can be classified in eight lipid categories of which not all are observed in human cells.[16,17] Lipids which are important building blocks in the human cell membrane belong to sterol lipids (e.g. cholesterol), glycerophospholipids (e.g. phosphatidyl-choline/lecithin), and sphingolipids (e.g. sphingomyelin). Sphingolipid- and cholesterol-rich lipid domains embedded in glycerophospholipids play central roles in the assembly of lipid rafts by forming an ordered and disordered liquid phase.[18,19] Lipid rafts are suggested to be involved in adipocyte physiology, cardiovascular disease, and carcinogenesis.[20] Glycerophospholipids are major components of biological membranes and play a key role in cell survival. Phosphatidyl-cholines (PCs), phosphatidylethanolamines (PEs), lyso-phosphatidylcholines (LPCs) and PE-plasmalogens (PEPLs) can be found in blood and the white matter of the central nervous system.[21,22] These glycerophospholipids play a crucial role in cell survival and cellular protection against reactive oxidant species (ROS) in peroxisomes through the PC, phosphatidylserine (PS) and PE rheostat.[23,24] Inflammation

is another pathway through which these glycerophospholipids play a role in common disease, in particular related to obesity, type 2 diabetes, and cardiovascular disease.[9] A wide range of disorders has been associated with altered circulating glycerophospholipids, in particular dyslipidemia, diabetes and coronary heart disease.[25–28] The sphingolipids comprise a complex range of lipids in which fatty acids are linked via amide bonds to a long-chain base or sphingoid. Sphingolipids also play essential roles as structural components of cell membranes and in cell signaling, and affect health and disease.[29,30]

In the following, we present the first comprehensive glycome- and lipidome-wide association study between 46 N-glycan and 183 lipid traits in the human blood of 2041 individuals from Croatia, Sweden, and Great Britain.

Experimental procedures

Study participants

We investigated three population-representative, pedigree-based, cross-sectional data collections from local populations in Croatia, Sweden, and Great Britain. All studies include a comprehensive collection of data on family structure, lifestyle, and blood samples for clinical chemistry, RNA and DNA analyses, and medical history. The studies received approval from the local ethics committees. All participants gave their written informed consent.[31] A brief description of each population is given below:

The VIS study was conducted in the villages of Vis and Komiza on the Dalmatian island of Vis, Croatia, between 2003 and 2004.[32–36] We analyzed 720 participants for whom relevant data was available, including 305 (42%) men and 415 (58%) women, aged between 18 and 93 years. The blood samples were collected under fasting conditions.

The NSPHS study (Northern Swedish Population Health Study) was conducted in the community of Karesuando in the subarctic region of the County of Norrbotten, Sweden, in 2006.[37,38] We studied 601 participants for whom relevant data was available including 281 (47%) men and 320 (53%) women, aged between 18 and 91 years. The blood samples were taken under non-fasting conditions to be able to model the effect of environmental covariates, e.g. dietary records, on blood metabolites.

The ORCADES study (Orkney Complex Disease Study) was conducted on the archipelago of Orkney, Scotland, Great Britain.[39,40] We examined 714 participants for whom relevant data was available including 331 (46%) men and 383 (54%) women, aged between 18 and 98 years.[32–36] The blood samples were collected under fasting conditions.

Analysis of N-glycans

Glycan release and labeling. The N-glycans from plasma sample (5 μl) proteins were released and labelled with 2-aminobenzamide (LudgerTag 2-AB labelling kit Ludger Ltd., Abingdon, UK), as described previously.[41] Labelled glycans were dried in vacuum centrifuge and re-dissolved in a known volume of water for further analysis.

Hydrophilic interaction high performance liquid chromatography (HILIC). Released glycans were subjected to hydrophilic

interaction high performance liquid chromatography (HILIC) on a 250 × 4.6 mm i.d. 5 μm particle packed TSKgel Amide 80 column (Tosoh Bioscience, Stuttgart, Germany) at 30 °C with 50 mM formic acid adjusted to pH 4.4 with ammonia solution as solvent A and acetonitrile as solvent B. 60 min runs were performed with fluorescence detector set with excitation and emission wavelengths of 330 and 420 nm, respectively. The system was calibrated using an external standard of hydrolyzed and 2-AB-labeled glucose oligomers from which the retention times for the individual glycans were converted to glucose units (GU).

The chromatograms obtained were all separated in the same manner to 16 chromatographic areas and 13 for desialylated glycans, regarding the peak resolutions and similarity of glycan structures present as described before. The amounts of glycans present in each area were expressed as a % of the total integrated chromatogram (amount of total glycan structures / total serum N-glycome).

Weak anion exchange (WAX)-HPLC. Glycans were separated according to the number of sialic acids by weak anion exchange HPLC. The analysis was performed using a Prozyme GlycoSep C 75 mm × 7.5 mm column (Prozyme, Leandro, CA, USA) at 30 °C with 20% (v/v) acetonitrile in water as solvent A and 0.1 M acetic acid adjusted to pH 7.0 with ammonia solution in 20% (v/v) acetonitrile as solvent B. Compounds were retained on the column according to their charge density, the higher charged compounds being retained the longest. A fetuin N-glycan standard was used for calibration.

Sialidase digestion. Aliquots of the 2AB-labeled glycan pool were dried down in 96-well PCR plates. To these, the following was added: 1 μl of 500 mM sodium acetate incubation buffer (pH 5.5), 1 μl (0.005 units) of ABS, *Arthrobacter ureafaciens* sialidase (releases α2–3,6,8 sialic acid, Prozyme) and H_2O to make up to 10 μl. This was incubated overnight (16–18 h) at 37 °C and then passed through AcroPrep™ 96-filter plates, 350 μL well, 10 K (Pall Corporation, Port Washington, NY, USA) before being applied to the HPLC.

Glycan structural features. Levels of glycans sharing the same structural features were approximated by adding the structures having the same characteristic from either HILIC, WAX or after sialidase treatment integrated glycan profiles.

Individual glycan structures present in each glycan group are shown in Supplementary Table S1, ESI.† Glycans were quantified from WAX profiles according the level of sialylation (monosialylated, disialylated, trisialylated, tetrasialylated). Glycan features were defined as: Core fucosylated glycans (FUC–C) = DG6/(DG5 + DG6)*100; Antennary fucosylated glycans (FUC-A) = DG7/(DG5 + DG7)*100; Biantennary glycans (BA) = DG1 + DG2 + DG3 + DG4 + DG5 + DG6 + DG7; Monosialylated biantennary glycans (BAMS) = (GP7 + GP8)/(DG5 + DG6 + DG7)*100; Disialylated biantennary glycans BADS = (GP9 + GP10 + GP11)/(DG5 + DG6 + DG7)*100; Triantennary glycans (TRIA) = DG8 + DG9 + DG10; Tetra-antennary glycans (TA) = DG11 + DG12 + DG13; Nongalacto-sylated glycans (G0) = DG1 + DG2; Monogalactosylated glycans (G1) = DG3 + DG4; Digalactosylated glycans (G2) = DG5 + DG6 + DG7; Trigalactosylated glycans

(G3) = GP12 + GP13 + GP14; Tetragalactosylated glycans (G4) = GP15 + GP16, Biantennary nongalactosylated glycan (A2) = (GP1 + DG1)/2 (Supplementary Table S2, ESI†).

Analysis of lipids

Total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides (TG) in the blood serum were quantified by enzymatic photometric assays using an ADVIA1650 clinical chemistry analyzer (Siemens Healthcare Diagnostics GmbH, Eschborn, Germany).

EDTA plasma samples were quantified upon lipid extraction by direct flow injection analysis.[42] A precursor ion scan of m/z 184 specific for phosphocholine containing lipids was used for phosphatidylcholine (PC), sphingomyelin (SM) and lysophosphatidylcholine (LPC).[2]

Glycerophospholipids and sphingolipids in the blood plasma were quantified by electrospray ionization tandem mass spectrometry (ESI-MS/MS) in positive ion mode as described previously.[42–45] A neutral loss scan of m/z 141 was used for phosphatidylethanolamine (PE)[27] and PE-based plasmalogens (PE-pls) which were analyzed according to principles described elsewhere.[46] Fragment ions of m/z 364, 380 and 382 were used for PE p16:0, p18:1 and p18:0 species, respectively. Ceramide (CER) was analyzed using a fragment ion of m/z 264.[45]

Quantification was achieved by calibration lines generated by addition of naturally occurring lipid species to plasma and internal standards belonging to the same lipid class (PC 14:0/ 14:0, PC 22:0/22:0, PE 14:0/14:0, PE 20:0/20:0, LPC 13:0, LPC 19:0). Calibration lines were generated for the following naturally occurring species: PC 34:1, 36:2, 38:4, 40:0 and PC O 16:0/20:4; LPC 16:0, 18:1, 18:0; PE 34:1, 36:2, 38:4, 40:6. Deisotopic and data analysis for all lipid classes was performed by self-programmed Excel Macros according to the principles described previously.[42] Overall 183 lipids scores including 151 lipid species from eight lipid subclasses were used in the analysis (Supplementary Table S3, ESI†).

Statistical analysis

Association analysis. The studied cohorts are pedigree-based samples from rural populations. This means that many participants are related to each other to varying degrees, but much closer than a sample from an urban population. To obtain the correct statistical inference, we used linear mixed effects models to adjust the blood parameters for relatedness, sex, and age. We then used the residues to calculate Pearson r correlations and perform conventional t-tests for correlations. We created heatmap plots to visually present the correlational results.

Covariate selection. We adjusted the outcome for sex and age, to find generic interactive metabolic pathways independent of sex or age. We did not include other covariates, such as diet, BMI, smoking, or alcohol consumption, whose effects on some glycans or lipids have been demonstrated.[47,48] First, the inclusion of these covariates might remove important variation in glycans or lipids which can give valuable information on interactive pathways. Second, we wanted to use a generic

statistical model with covariates which are more or less relevant for all glycans or lipids.

Multiple testing. Overall we calculated associations between 46 glycan scores and 183 lipid scores in three different populations which resulted overall in 25 254 statistical tests. A simple Bonferroni-adjustment to control the global α error at a 0.05 level would have lead to a local α error threshold of $\alpha = 0.05/(46*183*3) \approx 2 \times 10^{-6}$. Therefore, we decided to choose a more powerful closed test procedure with a discovery and two replication steps to identify associations across populations and to avoid false positive findings. We used the VIS cohort as the discovery cohort because it was historically the first cohort for which all of the examined traits were available and additionally is the cohort with the largest amount of non-missing data. Since this cohort alone provided a large number of consistent, significant results, we did not use the remaining data sets for discovery, but for replication. Additionally, we adjusted the local type I error per analysis step for the number of performed tests. In the discovery analysis, we therefore adjusted for 46×183 tests, corresponding to a local α error threshold of $\alpha_{\text{discovery}} \approx 5 \times 10^{-6}$ since we used only the Croatian cohort. In the replication analysis we adjusted the local α error only for the number of significant associations in the discovery cohort, which we tried to replicate, leading to $\alpha_{\text{replication}} = 0.05/330 \approx 1 \times 10^{-4}$. We want to point out that these significance criteria are rather strict in our opinion for the following reasons: First, the examined omics traits are highly correlated which substantially reduces the number of independent tests. Second, the populations are substantially different regarding environmental and genetic factors which means that it is much more difficult to replicate an association in another cohort. Therefore, a failing replication can be caused either by a false positive finding in the discovery phase and a false negative finding in the replication phase, or by a true positive result in the discovery phase and a true negative result in the replication phase. In the last case, the discovery and replication cohort show a really different association pattern. Finally, we report all conducted statistical tests to enable the reader to evaluate the significance of the findings as recommended.[49]

Software. All analyses were performed with the statistical analysis system *R*, Version 2.7.2.[50]

Results

We found 330 significant glycan-lipid associations with Pearson's correlation coefficient r ranging from -0.27 to 0.34 and corresponding p-values between 1.45 E−19 and 4.83 E−6 in the VIS cohort (Fig. 1). 175 significant correlations ($r = [-0.29 - +0.41]$; $p = [3.02$ E−24; 4.87 E−06]) were observed in NSPHS (Fig. 2), and 165 significant correlations ($r = [-0.49; +0.49]$; $p = [6.40$ E−44; 4.71 E−06]) were found in ORCADES (Fig. 3).

Discovery in the Croatian Cohort (VIS)

Classical lipids. GP11, GP12, GP13, DG8, and DG10 glycan species showed mostly significant positive associations with TC, LDL-C, or TG except GP6 which correlated negatively ($r = [-0.21; 0.32]$, $p = [3.96$ E−18; 4.10 E−06]). Regarding glycan summary scores, BA, TRIA, Trisialo, and G3 were mostly positively correlated with TC, LDL-C, and TG apart from BA which associated negatively ($r = [-0.20; +0.23]$, $p = [3.78$ E−09; 8.67 E−07]). HDL-C did not show significant relationships with glycans.

Glycerophospholipids. Significant associations were observed for 142 PC-glycan associations, 17 LPC-glycan associations, 83 PE-glycan associations, and 7 PEPL-glycan associations ($r = [-0.27; +0.32]$; $p = [6.67$ E−18; 4.83 E−06]). PCs primarily associated with GP11, GP12, GP13, GP15, DG7, DG8, DG11 as directly measured glycan species and derived traits BA, TRIA, Disialo, Trisialo, and G3 ($r = [-0.27; +0.32]$; $p = [6.67$ E−18; 4.83 E−06]). LPCs mainly correlated with GP8, GP12, DG8, BAMS, Disialo ($r = [-0.18; +0.23]$; $p = [1.18$ E−09; 3.59 E−06]). PEs were primarily linked with GP4, GP12, GP13, DG10, BA, TRIA, Trisialo, G3 ($r = [-0.22; +0.26]$, $p = [1.99$ E−11; 4.75 E−06]). PEPLs were only associated with DG10 ($r = [0.18; 0.20]$, $p = [1.97$ E−07; 4.20 E−06]).

Sphingolipids. Significant associations were observed for 24 SPM-glycan and 42 CER-glycan associations. SPMs correlated mainly with DG8, DG10, BA, TRIA, and G3 ($r = [-0.20; +0.29]$, $p = [1.09$ E−14; 4.76 E−06]). CERs associated primarily with GP11, GP13, GP14, DG8, DG10, BA, TRIA, and G3 ($r = [-0.19 - 0.34]$, $p = [1.45$ E−19; 4.2 E−06]).

Replication in the Swedish Cohort (NSPHS)

Overall, 85 glycan-lipid associations discovered in VIS were replicated in NSPHS (Table 1). Classical lipids such as TC, LDL-C, and TG were correlated with GP12, DG8, DG10, Trisialo, and G3 ($r = [-0.16; +0.40]$; $p = [3.78$ E−23; 9.11 E−5]). Of all glycerophospholipids, PCs showed 39 significant correlations, especially with GP12 and GP13 which are trigalactosylated glycans. Additionally, 3 LPCs were associated with GP8 and DG8 ($r = [-0.19; -0.32]$; $p = [8.96$ E−15–9.38 E−5]). Of all sphingolipids, SPMs showed 13 significant correlations exclusively with DG10, and CERs with DG10 and GP11 ($r = [0.22; 0.32]$; $p = [3.24$ E−15; 9.23 E−8]).

Replication in the Scottish Cohort (ORCADES)

Overall, 31 glycan-lipid associations found in VIS were confirmed in ORCADES (Table 2). No correlations between classical lipids and glycans were replicated. Of all glycerophospholipids, PEs showed a total of 11, *i.e.*, 9 with PCs and 3 with LPCs, significant associations with GP12, GP15, DG8, BAMS, or TRIA ($r = [-0.20; +0.24]$, $p = [2.79$ E−10; 7.55 E−05]). Of all sphingolipids, CERs showed 7, but SPMs only one significant association exclusively with DG10 ($r = [0.15; 0.24]$, $p = [2.69$ E−10; 7.05 E−5]). A schematic representation of significantly associated molecular structures of *N*-glycans and lipids is provided in Fig. 4.

In total only 10 correlations between a subset of triantennary glycans (DG8, DG10) and various unsaturated phosphatidylcholine, saturated ceramide, and sphingomyelin
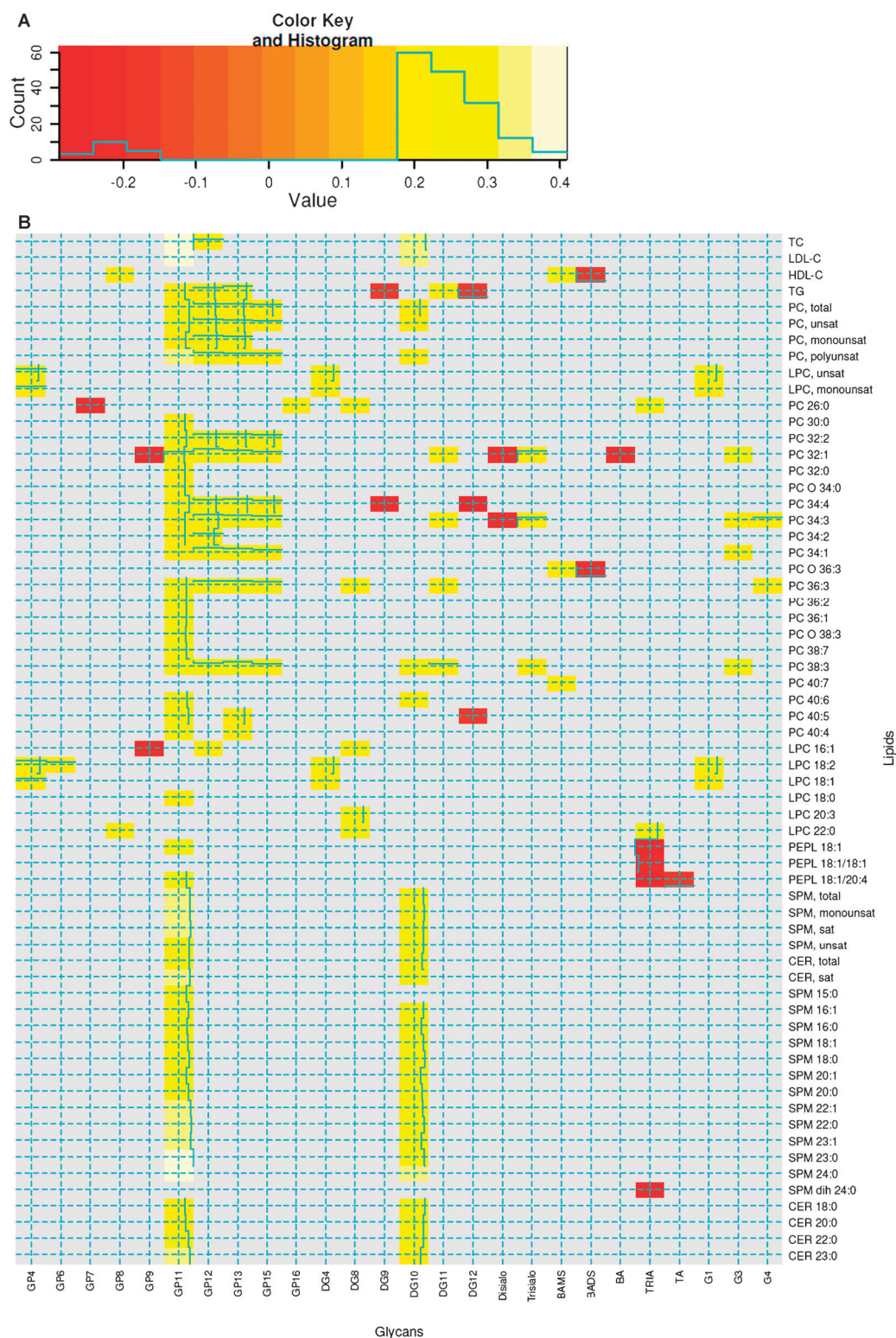
Fig. 1 Significant correlations between *N*-glycans and lipids in the VIS cohort (Croatia) under fasting conditions. Glycan and lipids which did not show any significant associations were omitted. Cells with non-significant correlations are displayed in grey. (A) Color key for the heatmap and a histogram of the overall distribution of Pearson *r* values. (B) Heatmap coding the Pearson *r* value in color.

lipids in VIS (r = [0.18; 0.34], p = [2.98 × 10^{-21}; 1.69 × 10^{-06}]) could be replicated in both NSPHS and ORCADES (Table 3). Association results for all glycan and all lipid parameters in all cohorts are available in Table S4, ESI.† Heatmaps displaying association results for all glycan and all lipid parameters in all cohorts are presented in Fig. S1–S10, ESI.†
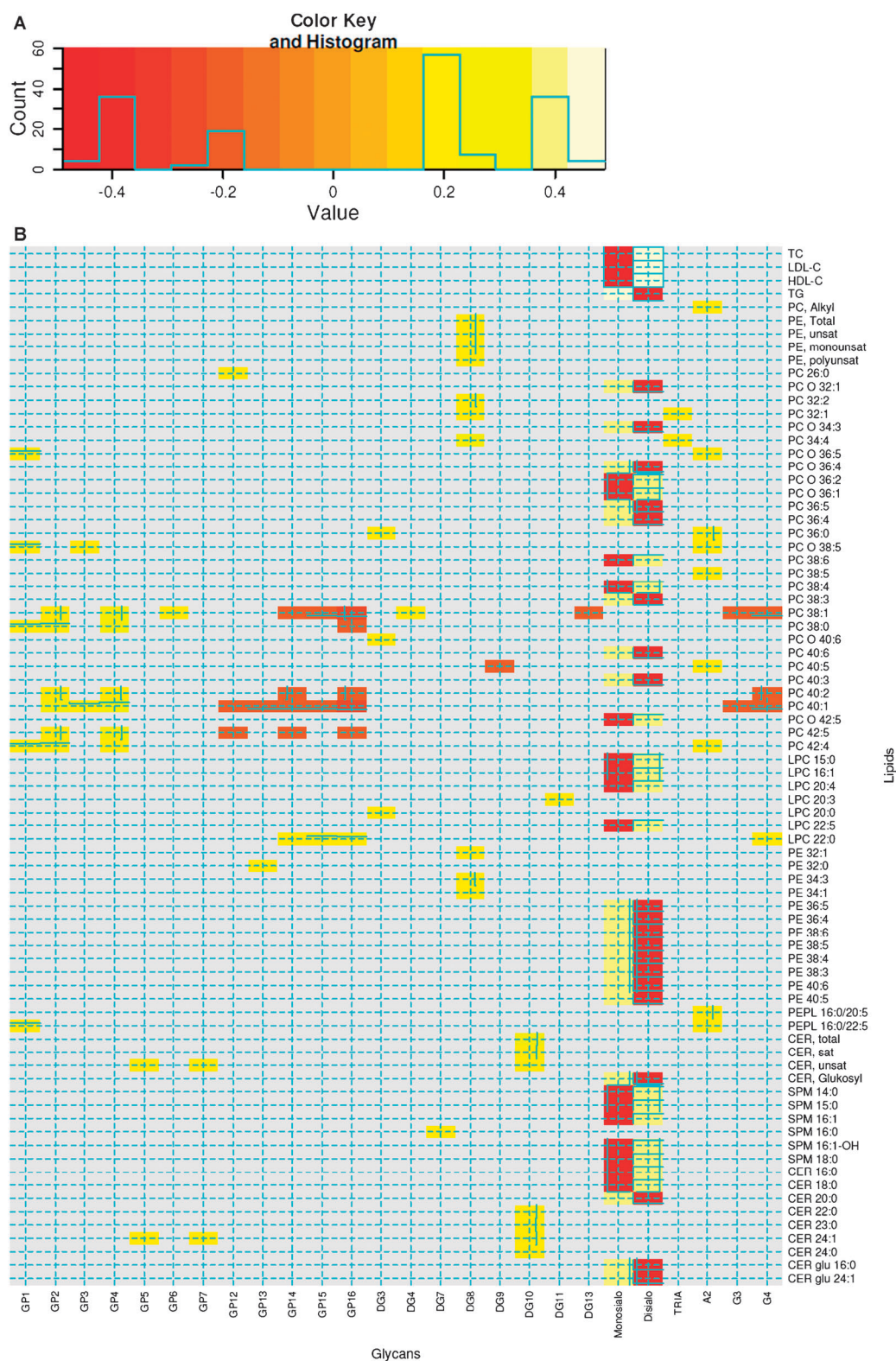
Fig. 2 Significant correlations between *N*-glycans and lipids in the NSPHS cohort (Sweden) under non-fasting conditions. Glycan and lipids which did not show any significant associations were omitted. Cells with non-significant correlations are displayed in grey. (A) Color key for the heatmap and a histogram of the overall distribution of Pearson *r* values. (B) Heatmap coding the Pearson *r* value in color.

Discussion

We conducted a comprehensive glycome-wide and lipidome-wide association study in three local populations in Croatia, Sweden, and Great Britain. We found strong associations between *N*-glycans and classical lipids, glycerophospholipids, and sphingolipids in each population and to a lesser degree also across populations. Bi- and triantennary glycans showed

Fig. 3 Significant correlations between *N*-glycans and lipids in the ORCADES cohort (Great Britain) under fasting conditions. Glycan and lipids which did not show any significant associations were omitted. Cells with non-significant correlations are displayed in grey. (A) Color key for the heatmap and a histogram of the overall distribution of Pearson *r* values. (B) Heatmap coding the Pearson *r* value in color.

a complementary pattern of glycan-lipid associations in Croatia, and partially in Sweden. A similar complementary pattern of mono- and disialylated glycans existed in Great Britain. The most prominent association between lipids and glycans was the repeated pattern of positive associations of polyunsaturated glycerolipids with triantennary (including trigalactosylated and trisialylated)

Table 1 Number of significant glycan and lipid associations in VIS (Croatia, fasting) which were replicated in NSPHS (Sweden, non-fasting)

LIPIDS	GLYCANS											
	DG10	GP12	GP13	G3	GP11	DG8	Trisialo	GP8	BA	Disialo	TRIA	Sum
PC	5	12	10	4	1	1	1		2	2	1	39
SPM	13											13
CER	6			3								9
TG	1			1		1	1					4
LPC						1		2				3
TC	1	1										2
LDL-C	1											1
Sum	27	13	10	5	4	3	2	2	2	2	1	71

Table 2 Number of significant glycan and lipid associations in VIS (Croatia, fasting) which were replicated in ORCADES (Great Britain, fasting)

LIPIDS	GLYCANS						
	DG8	TRIA	DG10	BAMS	GP15	GP12	Sum
PE	6	5					11
PC	4	3			1	1	9
CER			7				7
LPC	1			2			3
SPM			1				1
Sum	11	8	8	2	1	1	31

glycans, and negative correlations of the same lipids with biantennary (bigalactosylated and disialylated) glycans. For two prominent groups of triantennary glycans, DG8 and DG10, this association was replicated in all three populations. Although little is known about associations between glycans and lipids, we would like to discuss two potential interactive pathways between glycans and lipids which could account for these observations.

The number of antennas on N-glycans is determined by the activity of medial Golgi enzymes N-acetylglucosaminyl-transferases II, III, IV, and V that determine branching by the addition of N-acetlyglucosamine to core manoses.[51] If increased desaturation of glycerolphospholipids in medial Golgi is unfavorable for the activity of N-acetylglucosaminyl-transferases II, IV and V, this could result in a decrease of glycan branching and explain the observed associations.

Additionally, negative correlations between branched glycans and unsaturated fatty acids could be a sort of a compensatory mechanism. Changes in the degree of saturation of acyl chains in both PC and PE were reported to affect membrane domain formation and modulate phase separation from lipid raft molecules.[52,53] Lipid rafts are believed to modulate signaling events in a number of different ways according to the composition of the specific subpopulations. Polyunsaturated fatty acids and sterols possess a mutual aversion that drives the lateral segregation of phospholipids into highly disordered domains away from cholesterol. Triantennary glycans have the exact opposite effect on the formation of ordered lipid raft structures in the membrane. While polyunsaturated fatty acids increase disorder within the membrane bilayer, branched glycans increase order by promoting clustering of glycosylated membrane receptors by galectin-3.[54] Therefore, a possible explanation for the positive correlation between these two opposing functions is that individuals which have increased levels of polyunsaturated acids might compensate this by increasing glycan branching, and *vice versa*.

Strengths and weaknesses

We conducted the first large-scale, comprehensive glycome-wide and lipidome-wide association study in three rural, population-representative cohorts from Croatia, Sweden, and Great Britain. We examined a broad spectrum of glycomics and lipidomics traits which were quantified by state-of-the-art
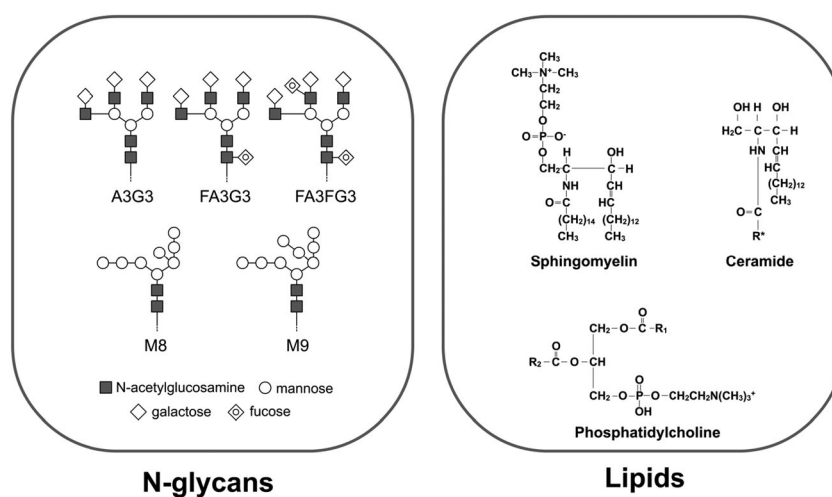
Fig. 4 A schematic representation of the glycan and lipid molecules that were found to be significantly associated in this study. *R* denotes saturated fatty acid.

Table 3 Glycans associated with lipids in VIS (Croatia) which were replicated in NSPHS (Sweden) and in ORCADES (Great Britain)

Glycan	Lipid	VIS		NSPHS		ORCADES	
		r	p*	r	p*	r	p*
DG8	PC 32 : 1	0.18	1.69 E−06	0.18	1.01 E−05	0.21	3.54 E−08
DG8	LPC 20 : 3	0.21	6.80 E−08	0.25	1.55 E−09	0.15	6.65 E−05
DG10	PC, unsat	0.21	1.78 E−08	0.21	3.55 E−07	0.14	3.36 E−04
DG10	PC, poly	0.23	2.35 E−09	0.21	2.53 E−07	0.15	1.27 E−04
DG10	CER, tot	0.34	1.82 E−19	0.26	5.73 E−10	0.24	2.69 E−10
DG10	CER, sat	0.34	1.45 E−19	0.28	1.11 E−11	0.23	1.95 E−09
DG10	SPM 18 : 1	0.28	5.46 E−14	0.29	2.13 E−12	0.14	2.86 E−04
DG10	SPM 22 : 0	0.28	1.39 E−13	0.29	1.20 E−12	0.15	7.05 E−05
DG10	CER 22 : 0	0.32	4.73 E−18	0.29	1.47 E−12	0.22	5.48 E−09
DG10	CER 23 : 0	0.31	5.50 E−17	0.22	9.23 E−08	0.21	4.07 E−08

Note: N(VIS) = 676, N(NSPHS) = 569, N(ORCADES) = 679 for all associations.

laboratory methods which have become available only recently. The glycan and lipid traits cover the most important species of these omics in human cells. These traits were quantified in a very large number of individuals in three different local populations in Europe. The remote location of these populations required substantially more financial and logistic resources than a comparable data collection in an urban population. The pedigree-based population structure made the application of sophisticated statistical models and high-performance computing servers necessary to obtain the correct statistical inference.

However, the scientific evidence from this study is also reduced for a number of reasons. First, the data has been collected using an epidemiological study design. Therefore, the association of two traits does not allow any conclusions about their casual relationship. Additionally, associations can be caused by other confounders which cause a statistical association without implying an interactive metabolic pathway. However, at the current, very early state of research on these omics traits, even these results can provide very important information and can point out potentially important metabolic pathways. Additionally, the specific patterns of associations found in each of the populations do not make an explanation by confounders very plausible.

Second, the populations show substantial differences regarding environmental and genetic factors, since they live in very different locations regarding geography, climate, and lifestyle. Therefore, the patterns of association between glycans and lipids can be truly different in each population, making it very difficult to replicate associations in another population. However, on the other side the associations found across all populations are probably extremely robust. Additionally, the differences between the populations will allow the study of environmental and genetic effects on omics traits in follow-up studies.

Third, we only measured glycan and lipid levels in the human blood, therefore, we cannot make reliable conclusions about the concentration or availability of the examined species in other tissues or body compartments. However, it is plausible to assume that the blood as the central transport medium will reflect bioavailability of these substances in the body.

Fourth, glycans were measured as relative levels compared to the overall glycome, lipids were measured as absolute levels. This should be kept in mind when interpreting the results as it can lead to methodological artifacts since an increase in the concentration of one glycan will affect the relative level of the glycan itself but also of all other glycans in the glycome. Therefore, it is unclear whether the reported complementary pattern of glycan-lipid associations has been caused by an increase of one glycan, a decrease of another one, or both, for example, if one glycan is metabolized to another glycan.

Fifth, we adjusted the glycan and lipid levels for sex and age. Therefore, associations between glycans and lipids which were caused by sex or age differences have been filtered out. One might argue to drop even those covariates to obtain the complete pattern of glycan-lipid associations. However, we chose this model to find generic potential metabolic pathways independent of sex and age effects.

Conclusions

Our study demonstrates that high-throughput studies of glycomics and lipidomics traits in large-scale population-based studies are feasible now. The results point towards generic interactive metabolic pathways between glycans and lipids across all populations, but also show specific associations for each population which can be caused by environmental and genetic differences.

Acknowledgements

References

1 B. F. Cravatt, G. M. Simon and J. R. Yates, *Nature*, 2007, **450**, 991–1000.

2 M. Scherer, G. Schmitz and G. Liebisch, *Clin. Chem.*, 2009, **55**, 1218–1222.

3 L. Royle, M. P. Campbell, C. M. Radcliffe, D. M. White, D. J. Harvey, J. L. Abrahams, Y. G. Kim, G. W. Henry, N. A. Shadick, M. E. Weinblatt, D. M. Lee, P. M. Rudd and R. A. Dwek, *Anal. Biochem.*, 2008, **376**, 1–12.

4 G. Cuccato, G. Della Gatta and D. di Bernardo, *Heredity*, 2009, **102**, 527–532.

5 L. Flintoft, *Nat. Rev. Genet.*, 2010, **11**, 3–3.

6 P. Geurts, A. Irrthum and L. Wehenkel, *Mol. BioSyst.*, 2009, **5**, 1593–1605.

7 K. X. Zhang and B. F. F. Ouellette, *Bioinformatics*, 2010, **26**, 529–535.

8 Y. S. Aulchenko, S. Ripatti, I. Lindqvist, D. Boomsma, I. M. Heid, P. P. Pramstaller, B. W. Penninx, A. C. Janssens, J. F. Wilson, T. Spector, N. G. Martin, N. L. Pedersen, K. O. Kyvik, J. Kaprio, A. Hofman, N. B. Freimer, M. R. Jarvelin, U. Gyllensten, H. Campbell, I. Rudan, A. Johansson, F. Marroni, C. Hayward, V. Vitart, I. Jonasson, C. Pattaro, A. Wright, N. Hastie, I. Pichler, A. A. Hicks, M. Falchi, G. Willemsen, J. J. Hottenga, E. J. de Geus, G. W. Montgomery, J. Whitfield, P. Magnusson, J. Saharinen, M. Perola, K. Silander, A. Isaacs, E. J. Sijbrands, A. G. Uitterlinden, J. C. Witteman, B. A. Oostra, P. Elliott, A. Ruokonen, C. Sabatti, C. Gieger, T. Meitinger, F. Kronenberg, A. Doring, H. E. Wichmann, J. H. Smit, M. I. McCarthy, C. M. van Duijn and L. Peltonen, *Nat. Genet.*, 2009, **41**, 47–55.

9 A. A. Hicks, P. P. Pramstaller, A. Johansson, V. Vitart, I. Rudan, P. Ugocsai, Y. Aulchenko, C. S. Franklin, G. Liebisch, J. Erdmann, I. Jonasson, I. V. Zorkoltseva, C. Pattaro, C. Hayward, A. Isaacs, C. Hengstenberg, S. Campbell, C. Gnewuch, A. C. W. Janssens, A. V. Kirichenko, I. R. König, F. Marroni, O. Polasek, A. Demirkan, I. Kolcic, C. Schwienbacher, W. Igl, Z. Biloglav, J. C. M. Witteman, I. Pichler, G. Zaboli, T. I. Axenovich, A. Peters, S. Schreiber, H. E. Wichmann, H. Schunkert, N. Hastie, B. A. Oostra, S. H. Wild, T. Meitinger, U. Gyllensten, C. M. van Duijn, J. F. Wilson, A. Wright, G. Schmitz and H. Campbell, *PLoS Genet.*, 2009, **5**, e1000672–e1000672.

10 R. T. Lee, G. Lauc and Y. C. Lee, *EMBO Rep.*, 2005, **6**, 1018–1022.

11 R. D. Cummings, *Mol. BioSyst.*, 2009, **5**, 1087–1104.

12 G. Lauc, I. Rudan, H. Campbell and P. M. Rudd, *Mol. BioSyst.*, 2010, **6**, 329–335.

13 K. W. Marek, I. K. Vijay and J. D. Marth, *Glycobiology*, 1999, **9**, 1263–1271.

14 J. D. Marth and P. K. Grewal, *Nat. Rev. Immunol.*, 2008, **8**, 874–887.

15 K. Ohtsubo and J. D. Marth, *Cell*, 2006, **126**, 855–867.

16 E. Fahy, S. Subramaniam, H. Brown, C. K. Glass, A. H. Merrill, 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, *J. Lipid Res.*, 2005, **46**, 839–861.

17 E. Fahy, 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, *J. Lipid Res.*, 2009, **50**(Suppl), S9-14–S19-14.

18 K. Jacobson, O. G. Mouritsen and R. G. Anderson, *Nat. Cell Biol.*, 2007, **9**, 7–14.

19 G. Schmitz and M. Grandl, *Curr. Opin. Clin. Nutr. Metab. Care*, 2008, **11**, 106–112.

20 V. Michel and M. Bakovic, *Biol. Cell*, 2007, **99**, 129–140.

21 G. van Meer, D. R. Voelker and G. W. Feigenson, *Nat. Rev. Mol. Cell Biol.*, 2008, **9**, 112–124.

22 M. Bakovic, M. D. Fullerton and V. Michel, *Biochem. Cell Biol.*, 2007, **85**, 283–300.

23 T. Illig, C. Gieger, G. Zhai, W. Römisch-Margl, R. Wang-Sattler, C. Prehn, E. Altmaier, G. Kastenmüller, B. S. Kato, H.-W. Mewes, T. Meitinger, M. H. de Angelis, F. Kronenberg, N. Soranzo, H. E. Wichmann, T. D. Spector, J. Adamski and K. Suhre, *Nat. Genet.*, 2010, **42**, 137–141.

24 G. Poli, G. Leonarduzzi, F. Biasi and E. Chiarpotto, *Curr. Med. Chem.*, 2004, **11**, 1163–1182.

25 G. Malerba, L. Schaeffer, L. Xumerle, N. Klopp, E. Trabetti, M. Biscuola, U. Cavallari, R. Galavotti, N. Martinelli, P. Guarini, D. Girelli, O. Olivieri, R. Corrocher, J. Heinrich, P. F. Pignatti and T. Illig, *Lipids*, 2008, **43**, 289–299.

26 Z. J. Zheng, A. R. Folsom, J. Ma, D. K. Arnett, P. G. McGovern and J. H. Eckfeldt, *Am. J. Epidemiol.*, 1999, **150**, 492–500.

27 B. Brugger, G. Erben, R. Sandhoff, F. T. Wieland and W. D. Lehmann, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 2339–2344.

28 A. M. Hodge, D. R. English, K. O'Dea, A. J. Sinclair, M. Makrides, R. A. Gibson and G. G. Giles, *Am. J. Clin. Nutr.*, 2007, **86**, 189–197.

29 S. T. Pruett, A. Bushnev, K. Hagedorn, M. Adiga, C. A. Haynes, M. C. Sullards, D. C. Liotta and A. H. Merrill, *J. Lipid Res.*, 2008, **49**, 1621–1639.

30 W. Zheng, J. Kollmeyer, H. Symolon, A. Momin, E. Munter, E. Wang, S. Kelly, J. C. Allegood, Y. Liu, Q. Peng, H. Ramaraju, M. C. Sullards, M. Cabot and A. H. Merrill, Jr., *Biochim. Biophys. Acta, Biomembr.*, 2006, **1758**, 1864–1884.

31 D. Mascalzoni, A. C. J. W. Janssens, A. Stewart, P. Pramstaller, U. Gyllensten, I. Rudan, C. M. van Duijn, J. F. Wilson, H. Campbell and R. McQuillan, *Eur. J. Hum. Genet.*, 2010, **18**, 296–302.

32 H. Campbell, A. D. Carothers, I. Rudan, C. Hayward, Z. Biloglav, L. Barac, M. Pericic, B. Janicijevic, N. Smolej-Narancic, O. Polasek, I. Kolcic, J. L. Weber, N. D. Hastie, P. Rudan and A. F. Wright, *Hum. Mol. Genet.*, 2007, **16**, 233–241.

33 O. Polasek, I. Kolcić, A. Smoljanović, D. Stojanović, M. Grgić, B. Ebling, M. Klarić, J. Milas and D. Puntarić, *Croat. Med. J.*, 2006, **47**, 649–655.

34 I. Rudan, Z. Biloglav, A. Vorko-Jovic, M. Kujundzic-Tiljak, R. Stevanovic, D. Ropac, D. Puntaric, B. Cucevic, B. Salzer and H. Campbell, *Croat. Med. J.*, 2006, **47**, 601–610.

35 I. Rudan, A. Marusic, S. Jankovic, K. Rotim, M. Boban, G. Lauc, I. Grkovic, Z. Dogas, T. Zemunik, Z. Vatavuk, G. Bencic, D. Rudan, R. Mulic, V. Krzelj, J. Terzic, D. Stojanovic, D. Puntaric, E. Bilic, D. Ropac, A. Vorko-Jovic, A. Znaor, R. Stevanovic, Z. Biloglav and O. Polasek, *Croat. Med. J.*, 2009, **50**, 4–6.

36 V. Vitart, Z. Biloglav, C. Hayward, B. Janicijevic, N. Smolej-Narancic, L. Barac, M. Pericic, I. M. Klaric, T. Skaric-Juric, M. Barbalic, O. Polasek, I. Kolcic, A. Carothers, P. Rudan,

N. Hastie, A. Wright, H. Campbell and I. Rudan, *Eur. J. Hum. Genet.*, 2006, **14**, 478–487.

37 W. Igl, Å. Johansson and U. Gyllensten, *Rur. Remote Health*, 2010, in press.

38 A. Johansson, F. Marroni, C. Hayward, C. S. Franklin, A. V. Kirichenko, I. Jonasson, A. A. Hicks, V. Vitart, A. Isaacs, T. Axenovich, S. Campbell, M. G. Dunlop, J. Floyd, N. Hastie, A. Hofman, S. Knott, I. Kolcic, I. Pichler, O. Polasek, F. Rivadeneira, A. Tenesa, A. G. Uitterlinden, S. H. Wild, I. V. Zorkoltseva, T. Meitinger, J. F. Wilson, I. Rudan, H. Campbell, C. Pattaro, P. Pramstaller, B. A. Oostra, A. F. Wright, C. M. van Duijn, Y. S. Aulchenko and U. Gyllensten, *Hum. Mol. Genet.*, 2009, **18**, 373–380.

39 R. McQuillan, A.-L. Leutenegger, R. Abdel-Rahman, C. S. Franklin, M. Pericic, L. Barac-Lauc, N. Smolej-Narancic, B. Janicijevic, O. Polasek, A. Tenesa, A. K. Macleod, S. M. Farrington, P. Rudan, C. Hayward, V. Vitart, I. Rudan, S. H. Wild, M. G. Dunlop, A. F. Wright, H. Campbell and J. F. Wilson, *Am. J. Hum. Genet.*, 2008, **83**, 359–372.

40 V. Vitart, A. D. Carothers, C. Hayward, P. Teague, N. D. Hastie, H. Campbell and A. F. Wright, *Am. J. Hum. Genet.*, 2005, **76**, 763–772.

41 A. Knezevic, O. Polasek, O. Gornik, I. Rudan, H. Campbell, C. Hayward, A. Wright, I. Kolcic, N. O'Donoghue, J. Bones, P. M. Rudd and G. Lauc, *J. Proteome Res.*, 2009, **8**, 694–701.

42 G. Liebisch, B. Lieser, J. Rathenberg, W. Drobnik and G. Schmitz, *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids*, 2004, **1686**, 108–117.

43 G. Liebisch, M. Binder, R. Schifferer, T. Langmann, B. Schulz and G. Schmitz, *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids*, 2006, **1761**, 121–128.

44 G. Liebisch, W. Drobnik, B. Lieser and G. Schmitz, *Clin. Chem.*, 2002, **48**, 2217–2224.

45 G. Liebisch, W. Drobnik, M. Reil, B. Trumbach, R. Arnecke, B. Olgemoller, A. Roscher and G. Schmitz, *J. Lipid Res.*, 1999, **40**, 1539–1546.

46 K. A. Zemski Berry and R. C. Murphy, *J. Am. Soc. Mass Spectrom.*, 2004, **15**, 1499–1508.

47 A. Knežević, O. Gornik, O. Polašek, M. Pučić, M. Novokmet, I. Redžić, P. M. Rudd, A. F. Wright, H. Campbell, I. Rudan and G. Lauc, *Glycobiology*, 2010, DOI: 10.1093/glycob/cwq1051, published online.

48 A. Knežević, O. Polašek, O. Gornik, I. Rudan, H. Campbell, C. Hayward, A. Wright, I. Kolčić, N. O'Donoghue, J. Bones, P. M. Rudd and G. Lauc, *J. Proteome Res.*, 2009, **8**, 694–701.

49 M. A. Proschan and M. A. Waclawiw, *Controlled Clin. Trials*, 2000, **21**, 527–539.

50 R Development Core Team, R, 2009, available from URL: http://www.r-project.org (accessed June 12, 2010).

51 *Handbook of glycosyltransferases and related genes*, ed. N. Taniguchi, K. Honke and M. Fukuda, Springer Verlag, Tokyo, 2002.

52 S. Raza Shaikh, A. C. Dumaual, D. LoCassio, R. A. Siddiqui and W. Stillwell, *Biochem.Biophys. Res. Commun.*, 2003, **311**, 793–796.

53 A. Hinderliter, R. L. Biltonen and P. F. Almeida, *Biochemistry*, 2004, **43**, 7102–7110.

54 J. W. Dennis, K. S. Lau, M. Demetriou and I. R. Nabi, *Traffic*, 2009, **10**, 1569–1578.