Supplementary tables, figures, and methods

25th November 2019

	Healthy volunteers ($n = 51$)	**T2DM only** ($n = 46$)	**CHD only** ($n = 25$)	**T2DM and CHD** ($n = 27$)
Female (%)	65	30	16	19
Age (years)	40 (12)	59 (8.4)	67 (9.7)	67 (5.7)
BMI (kg/m^2)	24 (5.1)	31 (5.7)	26 (4.2)	31 (4.1)
Waist circumference (cm)	87 (14)	110 (15)	100 (10)	110 (8.3)
Smoker (%)	9.8	20	20	15
Daily alcohol consumption (%)	7.8	13	20	15
Hypertension (%)	2	67	80	85
BP diastolic (mmHg)	80 (11)	87 (11)	76 (6.2)	82 (11)
BP systolic (mmHg)	120 (15)	130 (20)	120 (11)	130 (20)
Heart rate (bpm)	68 (11)	77 (9.3)	73 (12)	75 (11)
Hb (g/L)	140 (11)	150 (13)	140 (14)	140 (18)
HK (L/L)	0.42 (0.028)	0.44 (0.032)	0.42 (0.039)	0.42 (0.052)
Erythrocytes (T/l)	4.7 (0.43)	4.9 (0.46)	4.8 (0.59)	4.6 (0.49)
Leukocytes (G/l)	5.7 (1.8)	7.6 (2)	8.5 (1.8)	8.4 (2.8)
Thrombocytes (G/L)	250 (51)	240 (76)	230 (85)	230 (43)
CRP (mg/L)	1.2 (1.8)	2.4 (2)	5.2 (11)	5.8 (8.5)
ALT (U/L)	23 (13)	37 (30)	30 (20)	23 (8.5)
AST (U/L)	23 (6.7)	28 (14)	36 (22)	26 (8.6)
GGT (U/L)	19 (13)	51 (50)	39 (24)	41 (31)
eGFR (ml/min)	98 (14)	85 (18)	82 (27)	67 (22)
Glucose (mmol/L)	4.8 (0.5)	8.4 (2.7)	6.7 (2.5)	9.3 (3.9)
HbA1c IFCC (nmol/mol)	35 (3.6)	54 (11)	37 (4.7)	59 (13)
Cholesterol (mmol/L)	4.9 (1)	4.2 (1.1)	4.2 (1)	3.8 (0.97)
Triglycerides (mmol/L)	0.96 (0.64)	2 (1.5)	1.6 (0.68)	2.3 (2.2)
HDL-C (mmol/L)	1.8 (0.5)	1.2 (0.3)	1.2 (0.21)	1.2 (0.35)
LDL-C (mmol/L)	2.7 (0.95)	2 (0.73)	2.4 (0.89)	1.6 (0.6)
non-HDL-C (mmol/L)	3.1 (1.1)	3 (1.1)	3 (0.9)	2.6 (0.94)
ApoA-I (g/L)	1.6 (0.26)	1.4 (0.21)	1.3 (0.24)	1.4 (0.3)
ApoB (g/L)	0.83 (0.25)	0.82 (0.23)	0.85 (0.28)	0.78 (0.23)
Statins (%)	0	60.9	92	92.6
Insulin (%)	0	56.5	0	55.6
GLP1-related drugs (%)	0	56.5	0	25.9
Metformin (%)	0	76.1	0	66.7
Sulfonylureas (%)	0	8.7	0	18.5
Beta blockers (%)	0	17.4	84	70.4
Diuretics (%)	0	28.3	48	63
RAAS inhibitors (%)	1.96	63	76	88.9

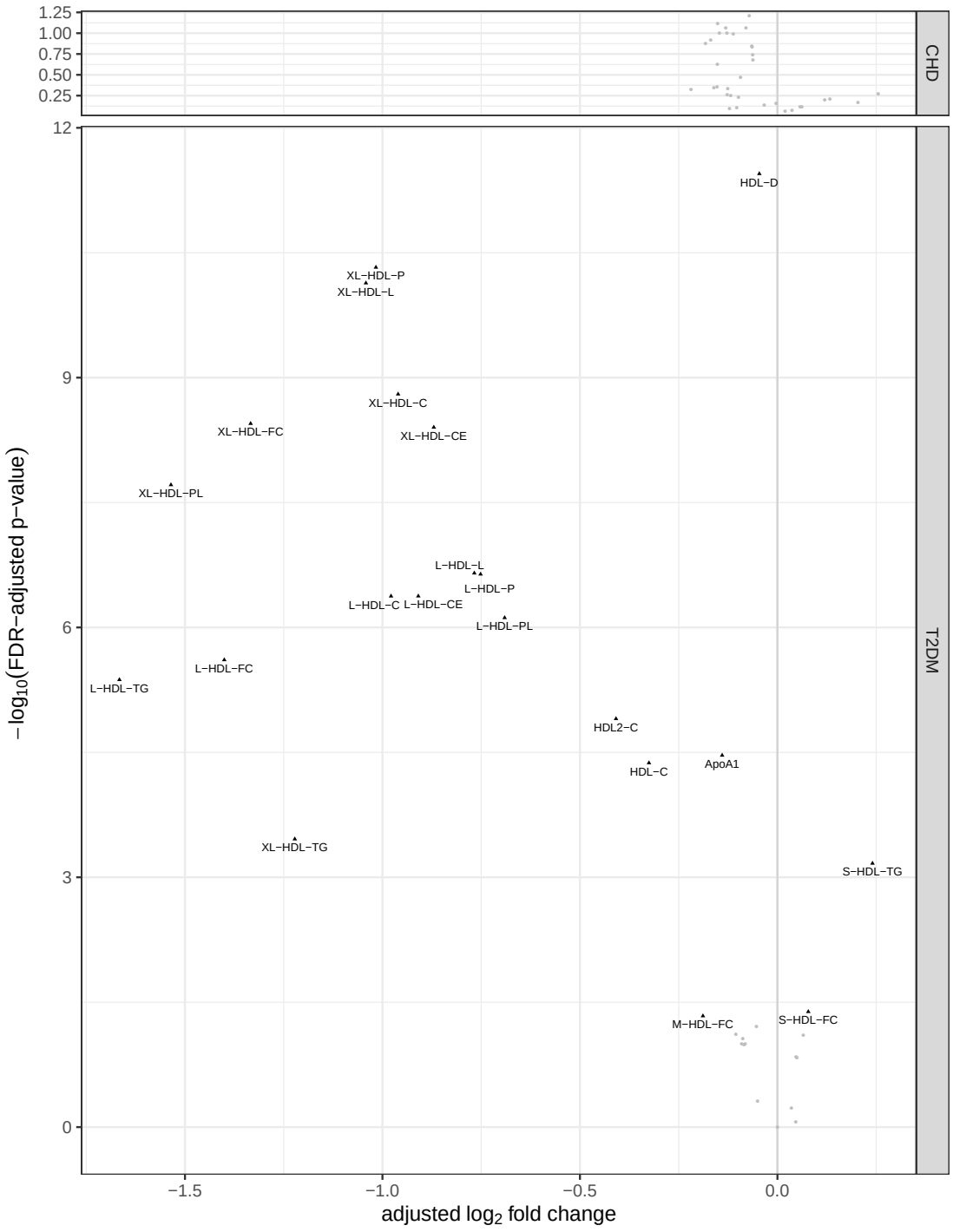
Supplementary Table S1: **Summary of clinical data for patients and healthy volunteers.** Average measurements are shown, with standard deviations in parenthesis. For dichotomous features, prevalences are shown as percentages. The bottom part of the table shows prevalence of drug use among study subjects.

HDL function	**Assay principle**	**Cells used**	**Calibrator/ reference**	**Inter-assay CV (%)**
Inhibition of endothelial cell apoptosis	ELISA of nucleosomes	Human aortic endothelial cells	rHDL	14.8
Cholesterol efflux by HDL	Fractional efflux of radiolabeled cholesterol	J774 macro-phages	reference plasma	9.7
Cholesterol efflux by apoB-free plasma				15.1
Inhibition of beta cell apoptosis	ELISA of nucleosomes	Ins1e cells	rHDL	14.3
Rescue of mitochondrial membrane potential	JC-1 fluorescence	C2C12 myotubes	rHDL	Max. 27.5
Mitochondrial respiration	Seahorse, Extracellular Flux Analyzer XFe96	Brown adipocytes differentiated from human multipotent adipose-derived stem cells	PBS, rHDL	Max. 15.0

Supplementary Table S2: **Characteristics of the bioassays for the recording of HDL functionality.** Inter-assay CV's were calculated from the data obtained from HDL isolated from pool plasma in eight series.

rHDL components	**low**	**medium**	**high**
rHDL + SM 42:2 or rHDL + SM 42:3: ApoA-I : DOPC : SM42:2 or SM42:3	1 : 96.6 : 3.33	1 : 90 : 10	1 : 96.6 : 33.3
rHDL + GPLD1: ApoA-I : DOPC : GPLD1	$1 : 100 : 10^{-5}$	$1 : 100 : 3 \times 10^{-4}$	$1 : 100 : 10^{-4}$
rHDL + apoF: ApoA-I : DOPC : apoF	$1 : 100 : 2 \times 10^{-5}$	$1 : 100 : 2 \times 10^{-4}$	$1 : 100 : 2 \times 10^{-3}$

Supplementary Table S3: **Relative molar composition of reconstituted HDL tested in the verification experiments.** "low", "medium", and "high" refer to the lowest, mean and highest concentrations of SM42:3, GPLD1, and apoF, respectively, measured in the native HDL samples. apoA-I and apoF = apolipoprotein A-I and F, respectively; DOPC = dioleyl phosphatidylcholine; GPLD1 = glycosylphosphatidylinositol-phospholipase D1; SM42:2 and SM42:3 = sphingomyelin d18:1/24:1 and sphingomyelin SM d18:2/24:1, respectively.



Supplementary Figure S1: **Volcano plot of associations between disease status and HDL features analysed by NMR spectroscopy of plasma.** Statistics based on linear regression models ($n = 149$) adjusting for hospital and gender (see §S4.1.2); t-tests of disease effects (contrasted with healthy volunteers) yielded p-values which were adjusted by the Benjamini–Hochberg procedure for controlling the false discovery rate. Only features with adjusted p-values smaller than 5% are labelled. Abbreviations of letters before "HDL": XL = very large, L = large, M = medium size, S = small. Abbreviations of letters after "HDL": C = cholesterol, CE = cholesterylesters, D = diameter, FC = free cholesterol, L = lipid, P = particle number, PL = phospholipids, TG = triglycerides.

The following lipoprotein particle features have FDR-adjusted p-values smaller than 5%, and are listed by increasing $\log_2$ fold change in parenthesis.

CHD

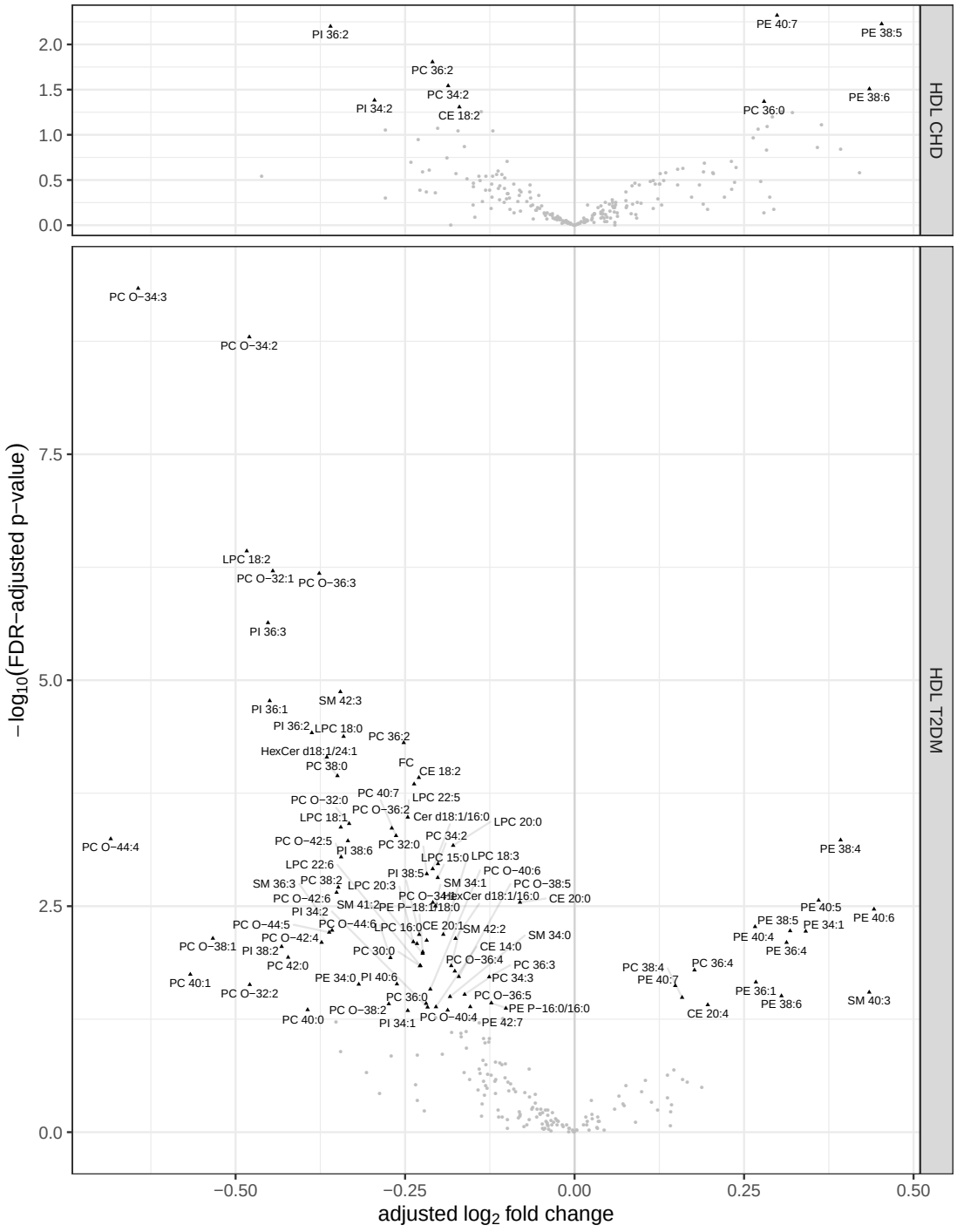
Negative $\log_2$ fold change (0 features)

Positive $\log_2$ fold change (0 features)

Diabetes

Negative $\log_2$ fold change (19 features): L-HDL-TG (-1.7), XL-HDL-PL (-1.5), L-HDL-FC (-1.4), XL-HDL-FC (-1.3), XL-HDL-TG (-1.2), XL-HDL-L (-1), XL-HDL-P (-1), L-HDL-C (-0.98), XL-HDL-C (-0.96), L-HDL-CE (-0.91), XL-HDL-CE (-0.87), L-HDL-L (-0.77), L-HDL-P (-0.75), L-HDL-PL (-0.69), HDL2-C (-0.41), HDL-C (-0.33), M-HDL-FC (-0.19), ApoA1 (-0.14), HDL-D (-0.046).

Positive $\log_2$ fold change (2 features): S-HDL-FC (0.078), S-HDL-TG (0.24).



Supplementary Figure S2: **Volcano plot of associations between disease status and lipids of HDL.** Statistics based on linear regression models ($n = 149$) adjusting for hospital, gender, HDL isolation date and rotor (see §S4.1.2); t-tests of disease effects (contrasted with healthy volunteers) yielded p-values which were adjusted by the Benjamini–Hochberg procedure for controlling the false discovery rate. Only lipid species with adjusted p-values smaller than 5% are labelled. For abbreviations see `https://www.lipidmaps.org/`

The following lipids have FDR-adjusted p-values smaller than 5%, and are listed by increasing $\log_2$ fold change in parenthesis.

CHD

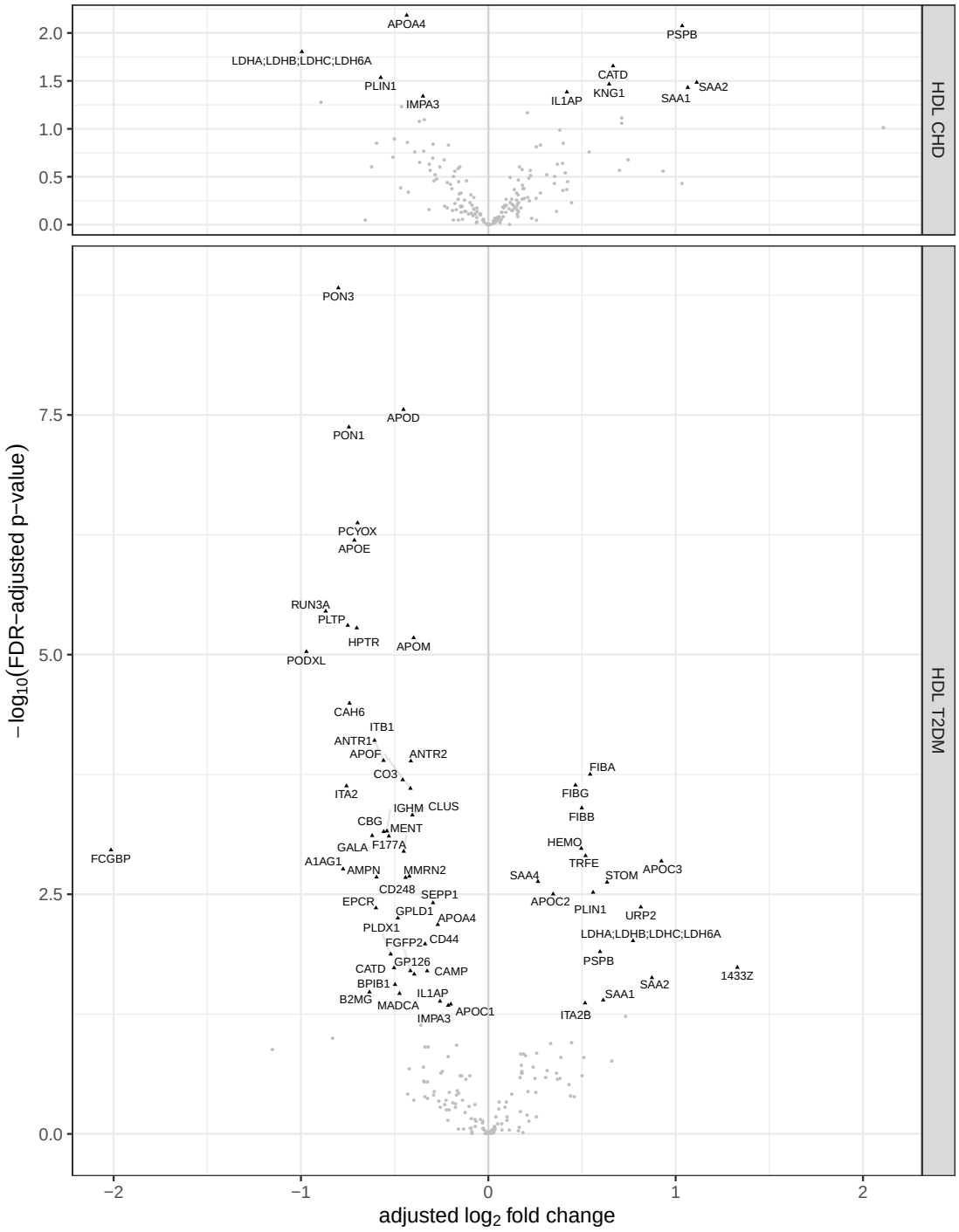
Negative $\log_2$ fold change (5 features): PI 36:2 (−0.36), PI 34:2 (−0.3), PC 36:2 (−0.21), PC 34:2 (−0.19), CE 18:2 (−0.17).

Positive $\log_2$ fold change (4 features): PC 36:0 (0.28), PE 40:7 (0.3), PE 38:6 (0.43), PE 38:5 (0.45).

Diabetes

Negative $\log_2$ fold change (71 features): PC O-44:4 (−0.68), PC O-34:3 (−0.64), PC 40:1 (−0.57), PC O-38:1 (−0.53), LPC 18:2 (−0.48), PC O-34:2 (−0.48), PC O-32:2 (−0.48), PI 36:3 (−0.45), PI 36:1 (−0.45), PC O-32:1 (−0.45), PI 38:2 (−0.43), PC 42:0 (−0.42), PC 40:0 (−0.39), PI 36:2 (−0.39), PC O-36:3 (−0.38), PC O-42:4 (−0.37), PC 38:0 (−0.37), PC O-44:5 (−0.36), SM 36:3 (−0.36), PC O-42:6 (−0.35), HexCer d18:1/24:1 (−0.35), PC 38:2 (−0.35), SM 42:3 (−0.35), LPC 18:1 (−0.34), PC O-42:5 (−0.34), LPC 18:0 (−0.34), PI 38:6 (−0.33), PC O-32:0 (−0.33), PE 34:0 (−0.32), PC O-38:2 (−0.27), PC O-44:6 (−0.27), PC O-36:2 (−0.27), PC 40:7 (−0.26), PI 40:6 (−0.26), PC 36:2 (−0.25), LPC 22:5 (−0.25), PI 34:1 (−0.25), LPC 20:3 (−0.24), CE 18:2 (−0.24), LPC 16:0 (−0.23), FC (−0.23), PI 38:5 (−0.23), PC 30:0 (−0.23), LPC 22:6 (−0.23), SM 41:2 (−0.22), PE P-18:1/18:0 (−0.22), PC 36:0 (−0.22), CE 20:1 (−0.22), PC 32:0 (−0.22), PI 34:2 (−0.22), PC O-40:6 (−0.21), Cer d18:1/16:0 (−0.21), PC O-34:1 (−0.21), SM 34:1 (−0.21), CE 14:0 (−0.2), LPC 15:0 (−0.2), PC 34:2 (−0.2), LPC 18:3 (−0.19), PC O-40:4 (−0.19), SM 34:0 (−0.18), SM 42:2 (−0.18), LPC 20:0 (−0.18), PC O-36:4 (−0.18), HexCer d18:1/16:0 (−0.18), PC O-38:5 (−0.17), PC 34:3 (−0.16), PC O-36:5 (−0.15), PC 36:3 (−0.13), PE P-16:0/16:0 (−0.12), PE 42:7 (−0.1), CE 20:0 (−0.081).

Positive $\log_2$ fold change (14 features): PE 40:7 (0.15), PC 38:4 (0.16), PC 36:4 (0.18), CE 20:4 (0.2), PE 40:4 (0.27), PE 36:1 (0.27), PE 38:6 (0.31), PE 36:4 (0.31), PE 38:5 (0.32), PE 34:1 (0.34), PE 40:5 (0.36), PE 38:4 (0.39), SM 40:3 (0.43), PE 40:6 (0.44).



Supplementary Figure S3: **Volcano plot of associations between disease status and proteins of HDL.** Statistics based on linear regression models ($n = 149$) adjusting for hospital, gender, HDL isolation date and rotor (see §S4.1.2); t-tests of disease effects (contrasted with healthy volunteers) yielded p-values which were adjusted by the Benjamini–Hochberg procedure for controlling the false discovery rate. Only proteins with adjusted p-values smaller than 5% are labelled. For protein abbreviations see `https://www.uniprot.org`

The following proteins have FDR-adjusted p-values smaller than 5%, and are listed by increasing $\log_2$ fold change in parenthesis.

CHD

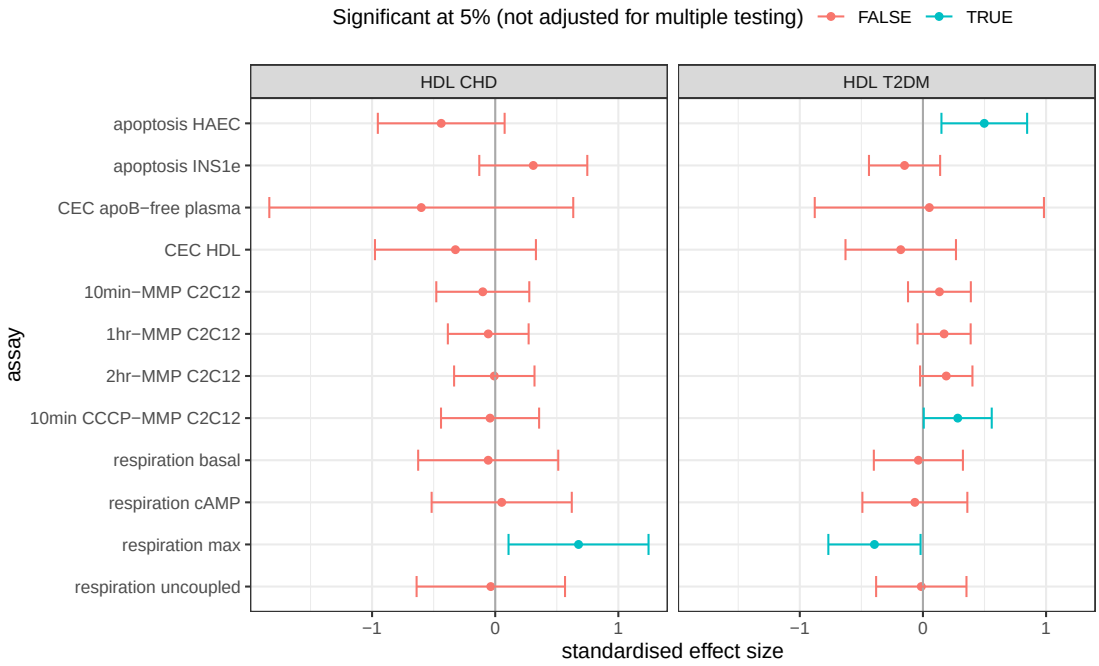
Negative $\log_2$ fold change (4 features): LDHA;LDHB;LDHC;LDH6A (−1), PLIN1 (−0.57), APOA4 (−0.44), IMPA3 (−0.35).

Positive $\log_2$ fold change (6 features): IL1AP (0.42), KNG1 (0.64), CATD (0.67), PSPB (1), SAA1 (1.1), SAA2 (1.1).

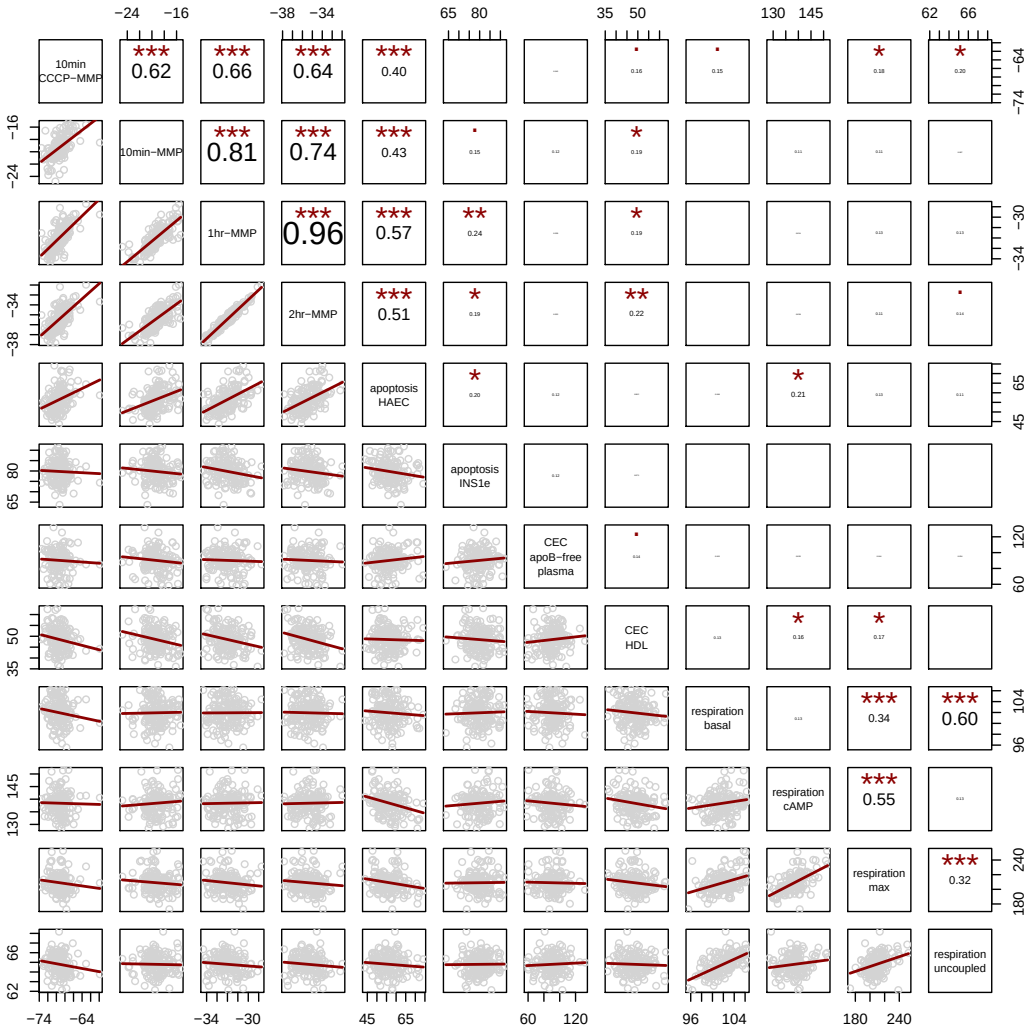
Diabetes

Negative $\log_2$ fold change (44 features): FCGBP (−2), PODXL (−0.97), RUN3A (−0.87), PON3 (−0.8), A1AG1 (−0.77), ITA2 (−0.76), PLTP (−0.75), PON1 (−0.74), CAH6 (−0.74), APOE (−0.71), HPTR (−0.7), PCYOX (−0.7), B2MG (−0.63), GALA (−0.62), ITB1 (−0.61), EPCR (−0.6), AMPN (−0.6), ANTR1 (−0.56), CBG (−0.56), IGHM (−0.54), F177A (−0.53), PLDX1 (−0.52), CATD (−0.5), BPIB1 (−0.5), GPLD1 (−0.48), MADCA (−0.47), CO3 (−0.46), APOD (−0.45), MENT (−0.45), CD248 (−0.44), MMRN2 (−0.42), APOF (−0.42), FGFP2 (−0.42), ANTR2 (−0.41), CLUS (−0.41), APOM (−0.4), GP126 (−0.4), CD44 (−0.34), CAMP (−0.33), SEPP1 (−0.3), APOA4 (−0.27), IL1AP (−0.26), IMPA3 (−0.21), APOC1 (−0.2).

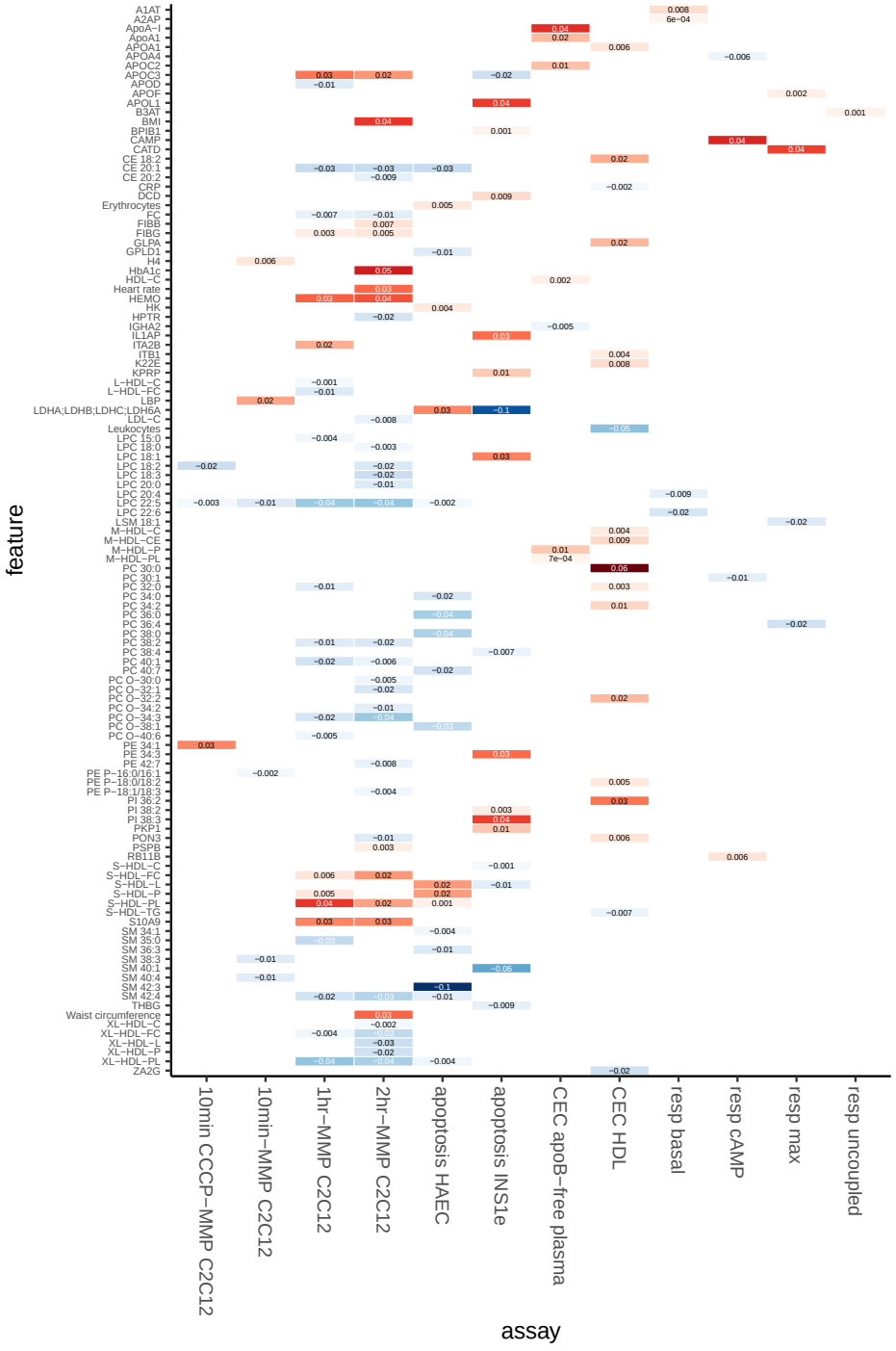
Positive $\log_2$ fold change (17 features): SAA4 (0.26), APOC2 (0.35), FIBG (0.47), HEMO (0.5), FIBB (0.5), ITA2B (0.52), TRFE (0.52), FIBA (0.54), PLIN1 (0.56), PSPB (0.6), SAA1 (0.61), STOM (0.63), LDHA/B/C/6A (0.77), URP2 (0.81), SAA2 (0.87), APOC3 (0.92), 1433Z (1.3).



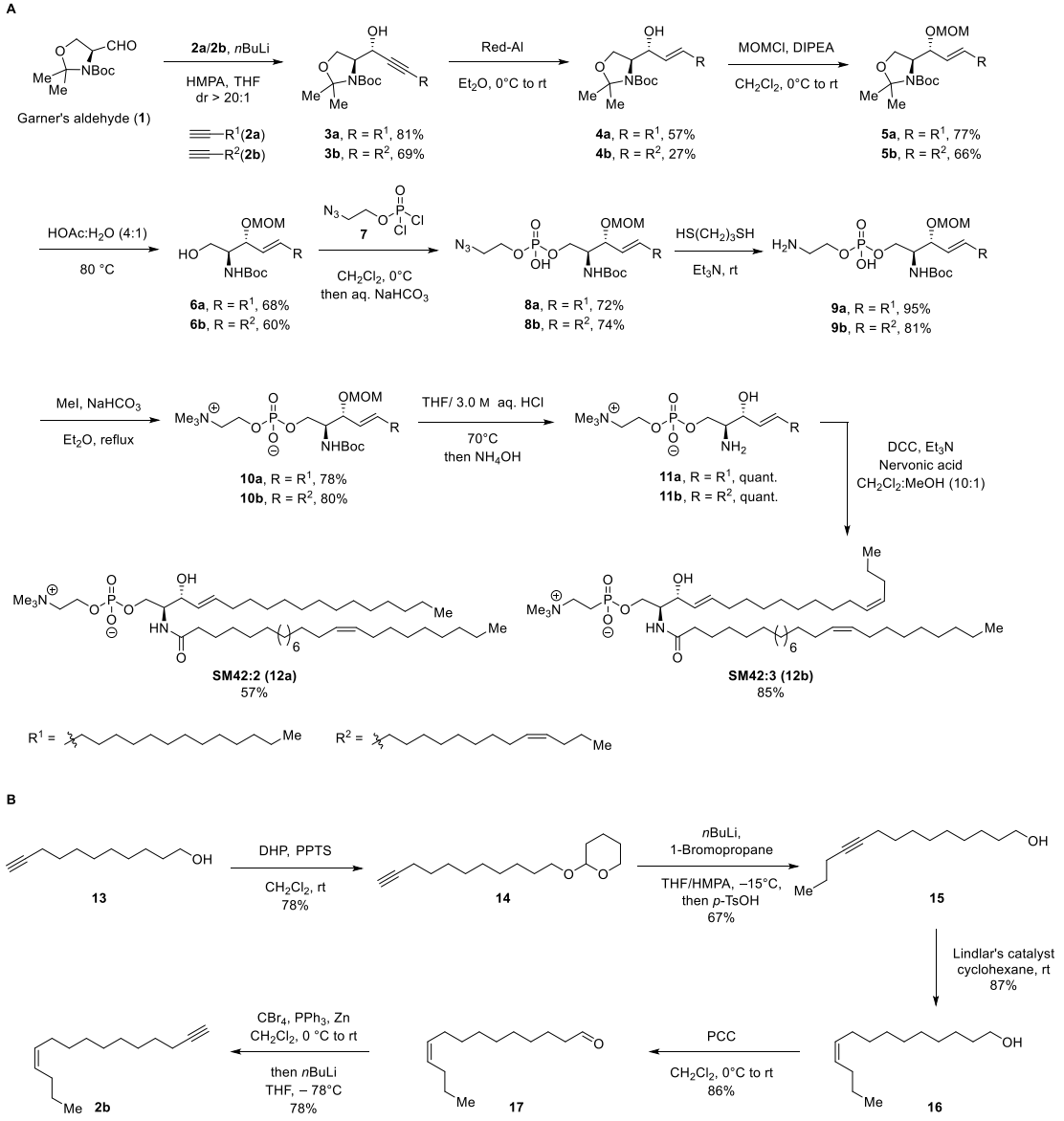
Supplementary Figure S4: **Standardised effect sizes of disease associations with HDL functions, estimated using mixed-effects models ($n = 149$). The figure shows 95% bootstrap confidence intervals for the standardised effects sizes of diseased HDL on the assayed cellular functions.** Numerical values are shown in Supplementary Table S4. If a confidence interval does not contain zero, the effect is significant at the 5% level, not adjusted for multiple comparisons. The analysis is adjusted for batch effects due to plate, HDL isolation date and centrifuge, as well as subject sex and sampling site. Abbreviations: CHD = coronary heart disease; HAEC = human aortic endothelial cells; INS1e = rat insulinoma cells; CEC apoB-free plasma = cholesterol efflux capacity of apoB-free plasma; CEC HDL = cholesterol efflux capacity of apoB-free plasma; 10 min-, 1hr-, 2hr-, 10 min CCCP-MMP C2C12 = mitochondrial membrane potential in C2C12 myotubes after 10 minutes, 1 hour, or 2 hours incubation with HDL or 10 minutes of incubation with the uncoupler Carbonyl cyanide m-chlorophenyl hydrazine. Basal, uncoupled, cAMP-stimulated and maximal mitochondrial respiration were recorded in human adipose-derived stem cells differentiated into brown adipocytes.



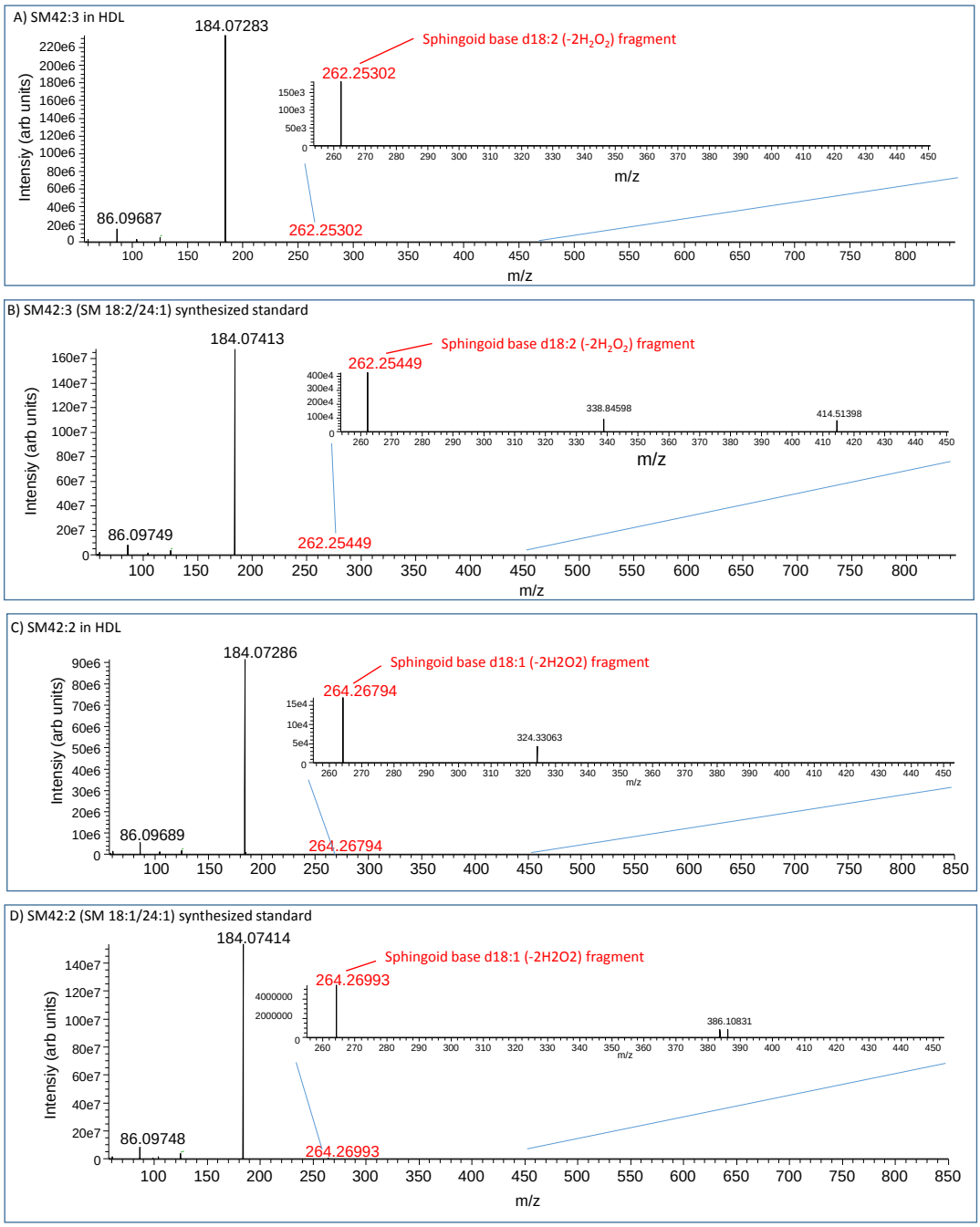
Supplementary Figure S5: **Pairwise correlations ($n = 149$) between HDL functions.** Printed above the diagonal are coefficients of Spearman correlation and asterisks for the level of statistical significance: *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$.



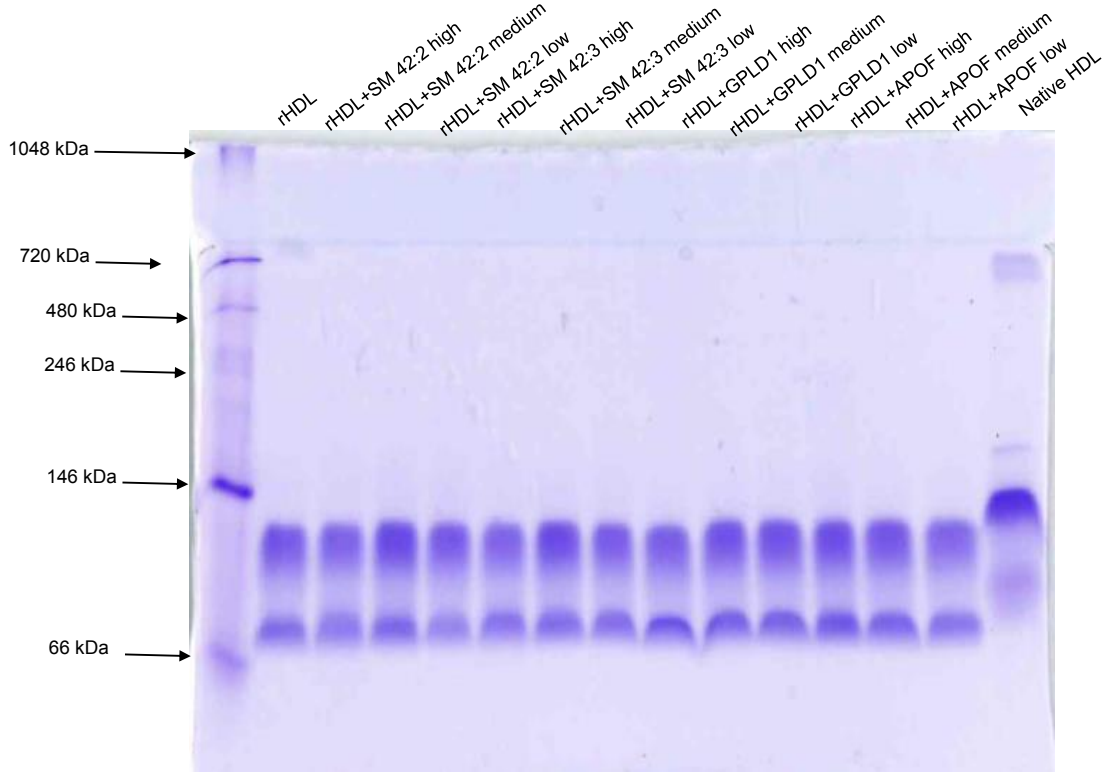
Supplementary Figure S6: **Gaussian graphical model estimating conditional dependencies between HDL function, structure, and subclasses, adjusted for clinical covariates.** For legibility, we show only the bipartite subgraph connecting HDL functions to other features. This is an alternative representation of the graph in Figure 4. The colour gradient indicates the sign and magnitude of the printed partial correlation, shrunken by the graphical lasso.



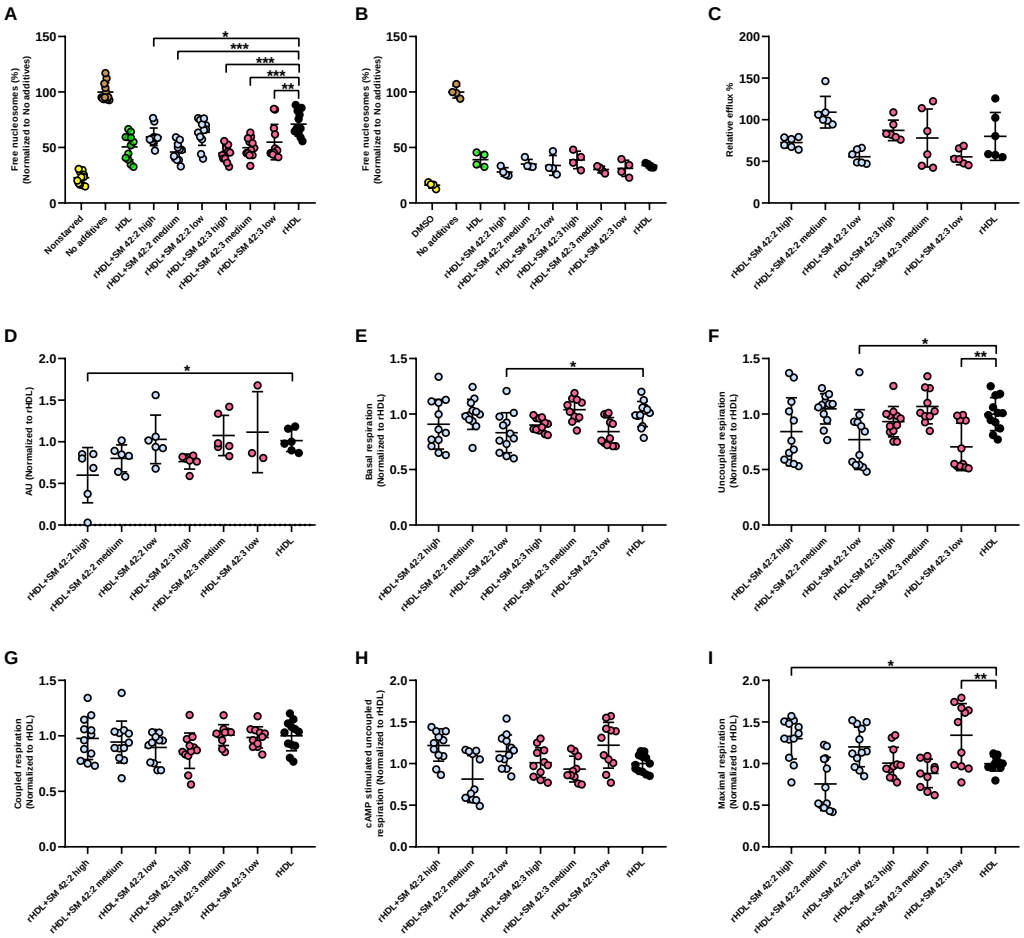
Supplementary Figure S7: **Schemata summarising the synthesis of sphingomyelin SM42:3 and SM 42:2.** For details and description of the molecules mentioned by numbers, see text in methods section. Part **A** shows the main steps of the synthesis. Part **B** shows the preparation of the alkyne 2b depicted in the upper line of figure **A**.



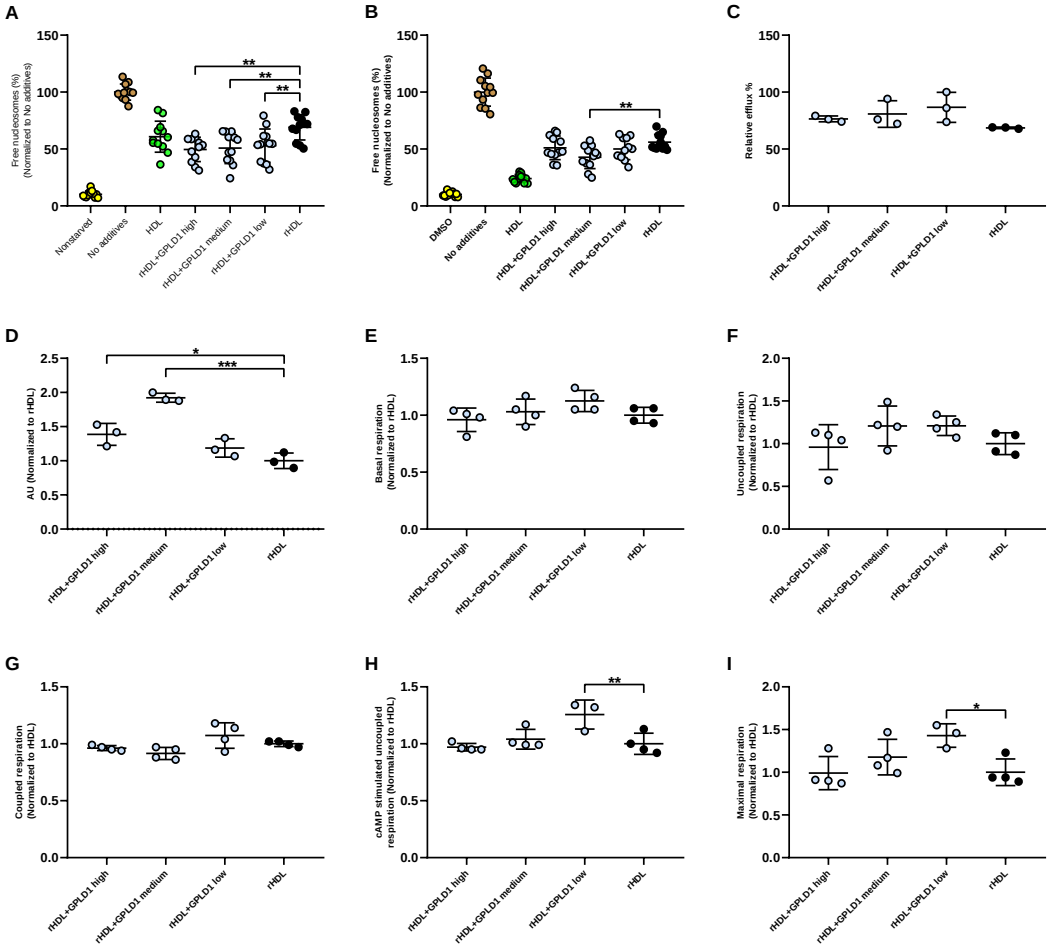
Supplementary Figure S8: **Comparison of LC-MS/MS spectra of SM42:3 (A) and SM42:2 (C) in HDL with synthetic sphingomyelins SM d18:2/24:1 (B) and SM d18:1/24:1 (D).**



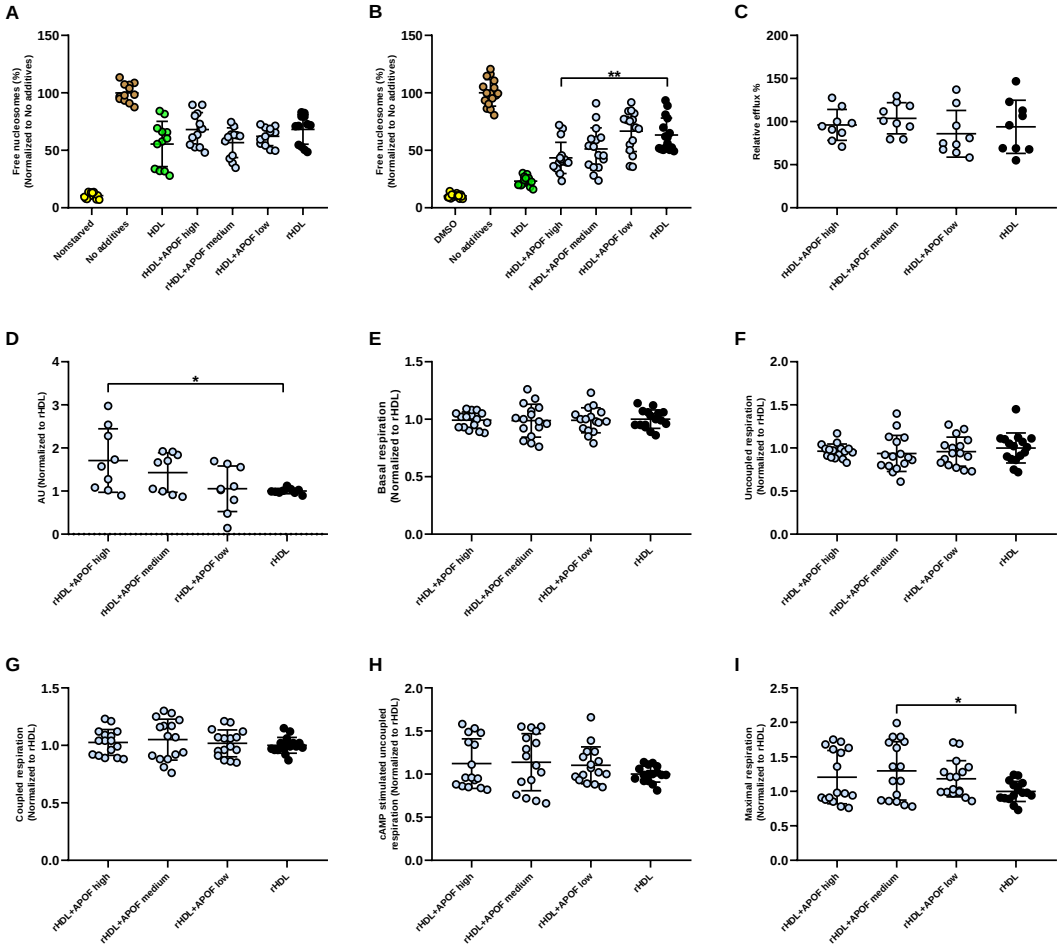
Supplementary Figure S9: **Polyacrylamide gradient gel electrophoresis of rHDL.** 1 µg of rHDL samples and native HDL were separated in 10% polyacrylamide gel in native conditions and stained with Coomassie brilliant blue.



Supplementary Figure S10: **Validation of SM 42:2 and SM 42:3 in the bioassays by the use of reconstitution of rHDL plus/minus SM 42:2 and SM 42:3.** SM 42:2 and SM 42:3 were reconstituted into rHDL with three different concentrations and tested in the bioassays. **(A)** Apoptosis in HAECs. **(B)** Apoptosis in INS1E cells. **(C)** Cholesterol efflux. **(D)** Mitochondrial potential of C2C12 myocytes. **(E)** Basal, **(F)** uncoupled, **(G)** coupled, **(H)** cAMP stimulated uncoupled and **(I)** maximal mitochondrial respiration in human brown hMADS cells. Data are represented as mean±SD at three dose levels, which were analyzed by one-way ANOVA coupled with Dunnett's test for multiple comparisons against rHDL. $***: P \leq 0.001$, $**: P \leq 0.01$, $*: P \leq 0.05$.



Supplementary Figure S11: **Validation of GPLD1 in the bioassays by the use of reconstitution of rHDL plus/minus GPLD1.** GPLD1 was reconstituted into rHDL with three different concentrations and tested in the bioassays. **(A)** Apoptosis in HAECs. **(B)** Apoptosis in INS1E cells. **(C)** Cholesterol efflux. **(D)** Mitochondrial potential of C2C12 myocytes. **(E)** Basal, **(F)** uncoupled, **(G)** coupled, **(H)** cAMP stimulated uncoupled and **(I)** maximal mitochondrial respiration in human brown hMADS cells. Data are represented as mean±SD at three dose levels, which were analyzed by one-way ANOVA coupled with Dunnett's test for multiple comparisons against rHDL. $***: P \leq 0.001$, $**: P \leq 0.01$, $*: P \leq 0.05$.



Supplementary Figure S12: **Validation of APOF in the bioassays by the use of reconstitution of rHDL plus/minus APOF.** APOF was reconstituted into rHDL with three different concentrations and tested in the bioassays. **(A)** Apoptosis in HAECs. **(B)** Apoptosis in INS1E cells. **(C)** Cholesterol efflux. **(D)** Mitochondrial potential of C2C12 myocytes. **(E)** Basal, **(F)** uncoupled, **(G)** coupled, **(H)** cAMP stimulated uncoupled and **(I)** maximal mitochondrial respiration in human brown hMADS cells. Data are represented as mean±SD at three dose levels which were analyzed by one-way ANOVA coupled with Dunnett's test for multiple comparisons against rHDL. $***: P \leq 0.001$, $**: P \leq 0.01$, $*: P \leq 0.05$.

S1 Synthesis of sphingomyelins SM42:3 and SM42:2

As shown in Supplementary Figure S7**A**, the synthesis of SM42:2 (SM d18:1/24:1) and SM42:3 (SM d18:2/24:1) began with the diastereoselective addition of lithiated alkyne **2a/2b** to (*S*)-Garner's aldehyde (**1**) [1] to provide *erythro* alcohol **3a/3b** in high yields [2]. Reduction of the propargylic alcohol **3a/3b** with Red-Al and protection of the hydroxy group as MOM ether afforded **5a/5b**. 3-MOM ether of (2*S*,3*R*)-*N*-Boc-sphingosine **6a** and its unsaturated derivative **6b** were obtained by selective cleavage of the acetal with acetic acid in water at 80 °C. Next, phosphorylation was accomplished by addition of a solution of **6a/6b** in dry CH_2Cl_2 to a solution of 2-azidoethyl phosphorodichloridate [3] in the presence of pyridine, affording the phosphate ester **8a/8b** in good yield. Reduction of the azido group with 1,3-propanedithiol in the presence of dry triethylamine provided amine **9a/9b** in 95% and 81% yield, respectively [4]. Treatment with a large excess of MeI (28 equiv) in the presence of $NaHCO_3$ yielded phosphocholine **10a/10b**. Simultaneous deprotection of the *N*-Boc and *O*-MOM groups with 3.0 M aq. HCl followed by coupling with nervonic acid produced the final products, SM42:2 (**12a**) and SM42:3 (**12b**), in high overall yields.

Although alkyne **2a** is commercially available, the alkyne **2b** was prepared as shown in Supplementary Figure S7**B**. Accordingly, tetrahydropyranyl (THP) ether **14** of commercially available alcohol **13** was alkylated with propyl bromide to give alcohol **15** after THP removal [5]. Partial reduction over Lindlar's catalyst provided alkene **16**, and PCC oxidation of **16** led to aldehyde **17**, which was transformed into the desired alkyne **2b** using the Corey–Fuchs procedure.

S2 Proteomics

Peptides were separated by reversed-phase chromatography on a high-pressure liquid chromatography (HPLC) system equipped with an EASY-Spray column (RSLC C18, 2 µm, 50 cm x 75 µm) (Thermo Fisher Scientific), which was connected to a nano-flow HPLC combined with an autosampler (EASY-nLC 1200, Thermo Fisher Scientific). Peptides were eluted at 50 °C with a constant flow rate of 200 nl/min in a 30-min linear gradient from 5–32% buffer B (80% ACN, 0.1% FA) and 5 min 32-56%B. For data-dependent acquisition (DDA), MS1 mass spectra were acquired at 60,000 resolution (automatic gain control target value 3e6). Peptide ions in the mass range of 375–1,500 m/z were monitored. HCD MS/MS scans were performed at 15,000 resolution (automatic gain control target value 1e5). Normalized collision energy was set to 28%, dynamic exclusion was 15 s. For data-independent acquisition (DIA) [6], one MS1 scan (400 to 1210 m/z) was recorded, followed by 19 consecutive DIA segments acquired at 30,000 resolution (AGC target 1e6 and auto for injection time) with variable scan windows. Normalized collision energy was set to 28%. The spectra were recorded in profile mode.

DDA sample runs were used to generate a comprehensive spectral library of HDL particles from 72 pooled patient samples. DDA spectra were analyzed using ProteomeDiscoverer Version 2.1 (Thermo Fisher Scientific) with Sequest HT [7] and MS Amanda [8] as search engines. The identifications were filtered for 5% FDR on peptide and 1% FDR on protein level. Data files were searched against the human UniProt fasta database (state 19.06.2014). The spectral library was assembled in Spectronaut version 11 (Biognosys) [9] and contains 347 proteins groups and 3288 proteotypic peptides. DIA Data were analyzed using Spectronaut in default mode. For retention time calibration iRTs (Biognosys) were added to each sample. For quantitation, standard settings were employed, which included dynamic peak detection, automatic precision nonlinear iRT calibration, interference correction, local cross run normalization and MS2 Top 3 summed peptide quantitation. All results were filtered with a q-value of 0.01 (equal to an FDR of 1%) on the peptide level. The data matrix was filtered with a q-value percentile of 0.2 (all observations that pass q-value threshold of 0.01 in 20% of the cases were considered).

S3 Lipidomics

A fragment ion of *m/z* 184 was used for phosphatidylcholine (PC), sphingomyelin (SM) [10] and lysophosphatidylcholine (LPC) [11] Neutral loss fragments were used for the following lipid classes: Phosphatidylethanolamine (PE) and phosphatidylinositol (PI) with a loss of 141 and 277, respectively [12, 13]. PE-based plasmalogens (PE P) were analyzed according to the principles described by Zemski-Berry [14]. Sphingosine based ceramides (Cer) and hexosylceramides (HexCer) were analyzed using a fragment ion of *m/z* 264 [15]. Free cholesterol (FC) and cholesteryl ester (CE) were quantified using a fragment ion of *m/z* 369 after selective derivatization of FC [16]. Quantification was achieved using two non-naturally occurring internal standards (IS) for each lipid class (except for PI, SM was calculated using PC IS and PE-based plasmalogens were calculated using PE IS) and calibration lines generated by standard addition of a number of naturally occurring species to plasma. Deisotoping and data analysis for all lipid classes was performed by self programmed Excel Macros as described previously [10, 17]. Lipid species were annotated according to the recently published proposal for shorthand notation of lipid structures that are derived from mass spectrometry [18]. Glycerophospholipid species annotation was based on the assumption of even numbered carbon chains only. SM species annotation is based on the assumption that a sphingoid base with two hydroxyl groups is present.

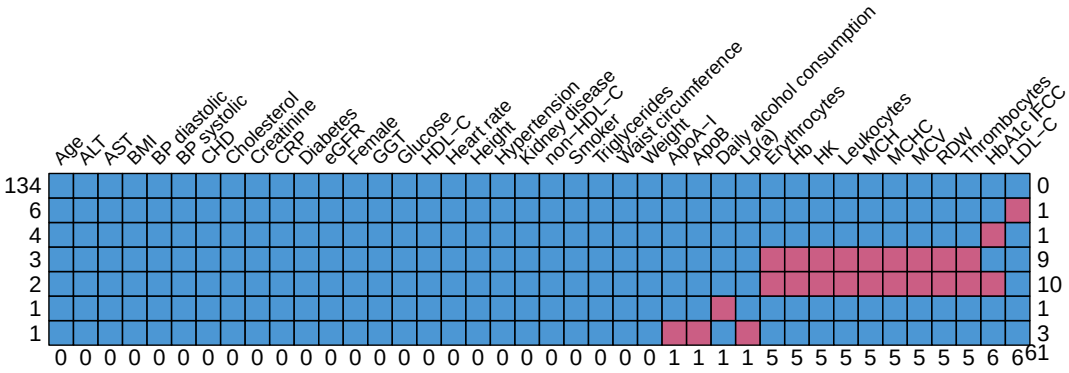
S4 Statistical methods

S4.1 Data preprocessing

For clinical variables, when a measurement was below the lower limit of quantification (LLOQ), we set them to $\text{LLOQ}/\sqrt{2}$; see [19].

S4.1.1 Imputation of clinical covariates

In the clinical data set, approximately 1% of values are missing at random (to the best of our knowledge), and we impute them using the R [20] package `mice` [21]. Supplementary Figure S13 summarises the missing data, where the top row corresponds to a fully observed profile, meaning that 134 patients have no missing data.



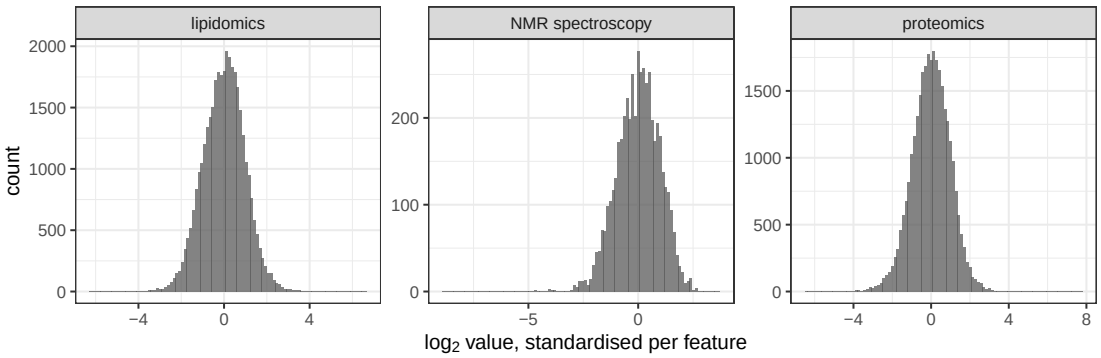
Supplementary Figure S13: Patterns of missing data, indicated by the burgundy red colour. Each row corresponds to a pattern of missing data (in terms of features), and the number of observations with the given pattern is displayed on the left. The number of patients with a given missingness pattern are displayed on the right. Below each column is the total number of patients with a missing value for the given feature. In the bottom right corner, the total number of missing data points (61) is displayed.

Using `mice`, we replace continuous missing data by their predicted values when linearly regressing the feature of interest on all other clinical covariates. Logistic regression is similarly used for binary variables. Ideally, we would run the imputation in many iterations adding noise, and evaluate how the downstream model fit varies. However, since our downstream analysis (see below) involves subsampling and cross-validation — making it non-deterministic and computationally intensive — we use the fitted values from single imputations.

S4.1.2 Normalisation of mass spectrometry and NMR spectroscopy data

Using mass spectrometry, 182 proteins and 227 lipid species were consistently quantified across the HDL samples in both isolation rounds. From plasma, abundances of HDL subclasses were measured using nuclear magnetic resonance (NMR) spectroscopy. For both technical and statistical reasons, it is advisable to logarithmically transform mass spectrometry data [22]. Indeed, in our datasets, log transforming and standardising yields bell-shaped distributions

(Supplementary Figure S14). However, some measurements are zero, which is problematic since $\log(0) = -\infty$ (in the



Supplementary Figure S14: Histograms of $\log_2$ transformed omics and NMR spectroscopy data, standardised per features to have sample mean 0 and variance 1. Original values of zero have been omitted.

limit). This is commonly addressed by adding an appropriate constant (pseudo-count) to all measurements before log transforming. For all features (proteins, lipid species, and HDL subclasses) separately, we estimate the lower limit of quantification (LLOQ) by the smallest non-zero measurement observed in the samples (provided that there exists a zero-valued measure of the given feature). When then add the estimated LLOQ to all measurements before applying the log transform. The magnitude of the pseudo-count affects the mean–variance relationship (scedasticity) of the log transformed omics variables; particularly for variables with a small mean. By assessing a wide and dense range of pseudo-counts, we found that adding an estimate of the LLOQ does not greatly compromise scedasticity.

Patients recruited and sampled at two different hospitals had their HDL isolated in two different rounds, on seven different dates, using two different centrifuge rotors. Principal component analyses (Supplementary Figure S15) clearly show an effect of the isolation round on both lipidomics and proteomics. However, they give no clear indication of effects from the specific isolation date, rotor or hospital. The NMR spectroscopy was performed on plasma, and is therefore not affected by HDL isolation procedures.

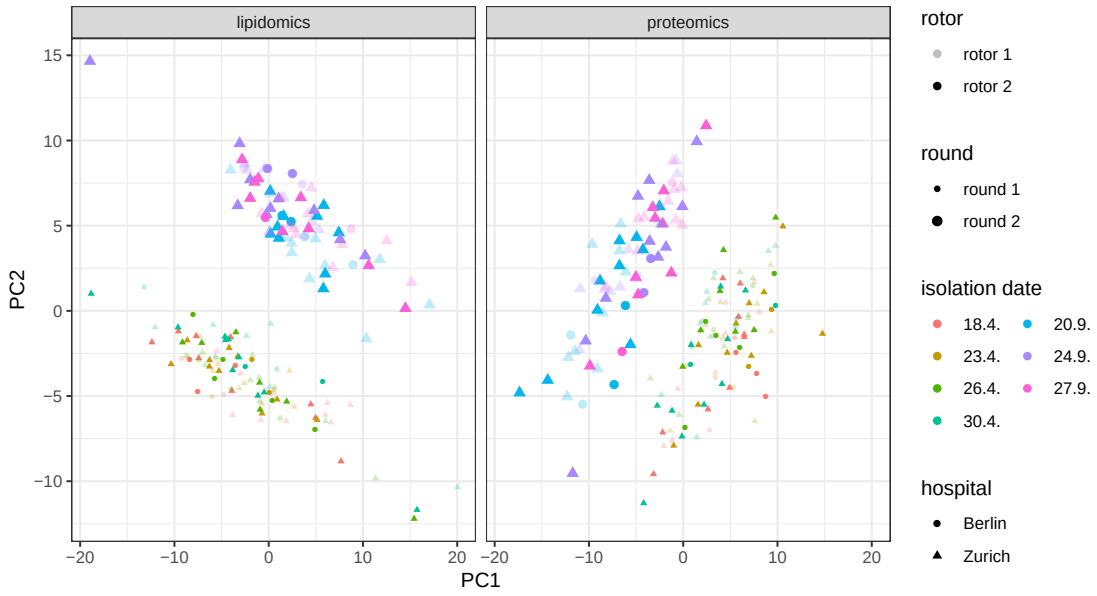
We use linear regression models to adjust the data for effects of hospital, gender, as well as HDL isolation date and rotor. (The isolation dates alias the rounds, so we need not adjust for the latter.) Naturally, we want to remove the effects of nuisance variables, but not the effects of disease conditions. The hospital effect is to a large extent confounded with the CHD signal, as most CHD patients were sampled in Berlin. Thus we include the disease effects in the model, to allow for patients with both diseases to help disentangle the hospital and CHD effects. Since the nuisance variables may influence each feature in a unique way, we fit separate linear models to all variables individually. Let Y_{ij} denote the $\log_2$-transformed value (after adding the estimated LLOQ) of the ith observation of the jth variable. Using indicator variables $\mathbb{1}[\cdot]$, we fit the linear model

$$Y_{ij} = \beta_{0j} + \beta_{1j}\mathbb{1}[\text{diabetes}_i = \texttt{TRUE}] + \beta_{2j}\mathbb{1}[\text{CHD}_i = \texttt{TRUE}] + \tag{1}$$
$$+ \beta_{3j}\mathbb{1}[\text{rotor}_i = 2] + \beta_{4j}\mathbb{1}[\text{hospital}_i = \text{Zurich}] + \tag{2}$$
$$+ \beta_{5j}\mathbb{1}[\text{gender}_i = \text{female}] + \vec{\beta}_{6j}^{\top}\vec{D}(i) + \varepsilon_{ij} \tag{3}$$

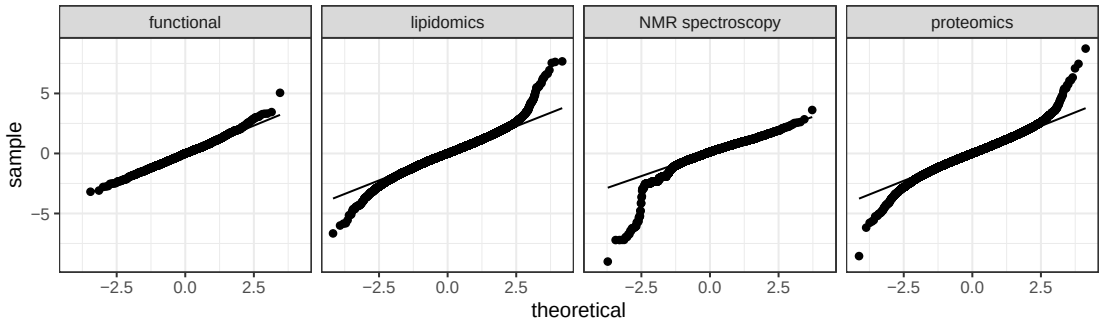
where ε_{ij} are independent and identically distributed, and $\vec{D}(i)$ is a binary vector with a 1 corresponding to the HDL isolation date of sample i. The first level of each indicator variable has been absorbed into the intercept β_{0j}. We use

$$\hat{\beta}_{0j} + \hat{\beta}_{1j}\mathbb{1}[\text{diabetes}_i = \texttt{TRUE}] + \hat{\beta}_{2j}\mathbb{1}[\text{CHD}_i = \texttt{TRUE}] + \hat{\varepsilon}_{ij} \tag{4}$$

as a normalised value of Y_{ij} since the residual $\hat{\varepsilon}_{ij}$ contains all information from Y_{ij} after linearly regressing out disease and nuisance variables [23]. We retain the disease effects for downstream analysis. After normalisation, the standardised distributions have heavier tails (larger kurtosis) than a normal distribution (Supplementary Figure S16).



Supplementary Figure S15: Principal component analyses of omics data after adding a pseudo-count and log transforming. Each variable was centred and scaled before PCA. There is clear separation based on HDL isolation round (symbol size). However, within each round, the isolation date (colour) does not appear to have an effect. Neither the hospital (shape) nor the rotor (transparency) discriminates between points.



Supplementary Figure S16: Normal Q–Q plots of the data adjusted for hospital, gender, and, when applicable, HDL isolation date and rotor.

S4.1.3 Normalisation of functional bioassays

Bioassays were used to measure functional properties of HDL isolated from samples of patients and health volunteers. Generally, samples were assayed using 96-well plates, in 2–5 technical replicates (different wells) per HDL sample. However, the experimental designs varied between bioassays. In some, all technical replicates were put on the same plate, whereas in others, replicates were distributed across plates.

We used the R package `lme4` [24] to fit mixed-effects models accounting for variation between plates, whilst utilising all technical replicates for estimation. In addition to the fixed effects in Equations (1)–(3), we fitted fixed effects for the plates, and a random effect σ_{ij} for sample i in bioassay j. In some cases the HDL isolation dates or rotor aliased plates, and redundant levels were dropped. The normalised measurement of sample i in bioassay j was computed as in Equation (4), though with $\hat{\sigma}_{ij}$ in lieu of $\hat{\varepsilon}_{ij}$, since the latter, in the mixed-effects models, captures technical rather than biological variation. The normalised data is approximately normally distributed (Supplementary Figure S16).

S4.2 Logistic regression of disease status

For a given disease status (diabetes, CHD, or both), we perform linear logistic regression of disease versus healthy, with all features (clinical, structural, and functional) as explanatory variables. We do this for each disease status separately, for the following reasons. The multivariate regression problem has the same least squares solution as the solutions to separate regressions, even if the response variables are correlated [25] (section 3.2.4). One way to bridge the two regressions is to regularise them jointly, using the group lasso on each covariate across response variables [26] (section 4.3). The same approach is taken by other methods of multitask learning [27]. However, we expect that diabetes and CHD may have different predictors, in which case it would be more appropriate to regularise them separately.

Let Y be a binary random variable indicating disease. We use the logit link to model

$$g(x) := \log \frac{P(Y = 1 \mid X = x)}{P(Y = 0 \mid X = x)} = \beta_0 + \beta^\top x$$

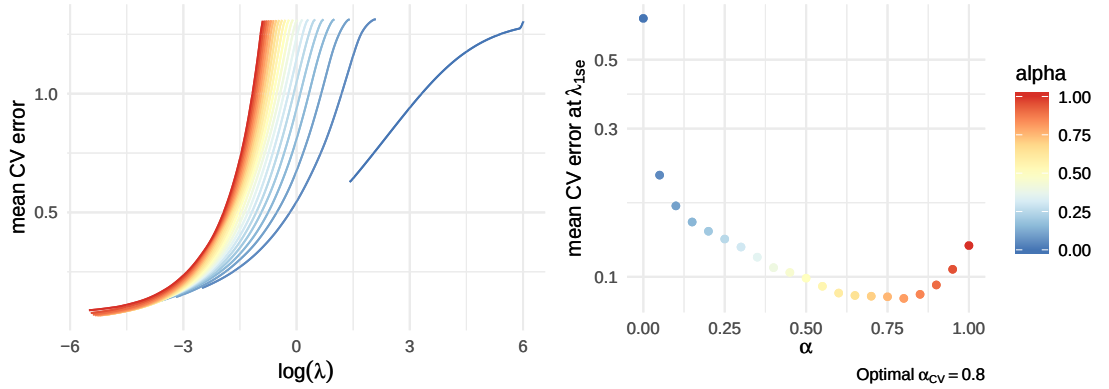
where X is a random vector of p features. Since in our case $p > n$, regularisation is required to make the model identifiable. Given n observations $(y_i, x^{(i)})_{i=1}^n$, the negative log likelihood with elastic net regularisation is

$$-\frac{1}{n}\sum_{i=1}^{n}\left[y_i(\beta_0 + \beta^\top x^{(i)}) - \log(1 + e^{\beta_0 + \beta^\top x^{(i)}})\right] + \lambda\left[\frac{1}{2}(1-\alpha)\|\beta\|_2^2 + \alpha\|\beta\|_1\right] \tag{5}$$

which we aim to minimise with respect to $(\beta_0, \beta) \in \mathbb{R}^{p+1}$. We use the elastic net rather than the lasso [28], because when a group of variables are correlated, the lasso tends to randomly select one predictor from the group, whereas the

elastic net shrinks correlated variables together and selects the entire group [29]. However, we use cross-validation to tune the elastic net mixing parameter α in formula (5), which would yield the lasso if $\alpha = 1$.

We use `glmnet` [30] to perform elastic net regression, and `glmnetUtils` [31] to run cross-validation of both parameters α and λ simultaneously. The values of the parameter α range from 0 to 1, in increments of 0.01. For a given value of α, following the default settings of `glmnet`, the range of λ is chosen by setting the lower bound to $\lambda_{\text{upper}} \cdot 10^{-2}$, where the upper bound λ_{upper} is determined (from the data) as the smallest value for which all regression coefficients are zero. Then 100 parameter values of λ are spaced evenly on a logarithmic scale between the lower and upper bound of lambda. For each point, the mean cross-validation (CV) error (binomial deviance) along with its standard error is computed. For a given α, let λ_{min} denote the largest value of λ which attains the minimal mean CV error cvm_{min} with standard error se_{min}. Let $\lambda_{1\text{se}}$ denote the largest λ whose mean CV error is smaller than $\text{cvm}_{\text{min}} + \text{se}_{\text{min}}$. The mean CV error at $\lambda_{1\text{se}}$ is denoted $\text{cvm}_{1\text{se}}$, and to emphasise that these quantities depend on α, we write $\lambda_{1\text{se}}(\alpha)$ and $\text{cvm}_{1\text{se}}(\alpha)$. The optimal α is $\alpha_{\text{CV}} := \arg\min_{\alpha}\{\text{cvm}_{1\text{se}}(\alpha)\}$. The optimal elastic net parameter pair $(\alpha, \lambda)_{\text{CV}}$ for formula (5) is therefore $\big(\alpha_{\text{CV}}, \lambda_{1\text{se}}(\alpha_{\text{CV}})\big)$. Supplementary Figure S17 shows the



Supplementary Figure S17: Simultaneous cross-validation of the elastic net parameters α and λ in the logistic regression of diabetes and CHD status. **Left:** mean CV error curves for a sweep of $\alpha \in [0,1]$. **Right:** mean CV error at $\lambda_{1\text{se}}(\alpha)$ for $\alpha \in [0,1]$.

result of simultaneous cross-validation of α and λ in the logistic regression of patients with both diabetes and CHD vs healthy volunteers. The optimal parameter pair $(\alpha, \lambda)_{\text{CV}} = (0.85, 1.16 \cdot 10^{-2})$.

We applied the above methodology in three separate logistic regression models contrasting a disease condition with healthy volunteers. The disease conditions were defined as having (1) only diabetes, (2) only CHD, and (3) both diabetes and CHD. The coefficients included in the models with elastic net parameters $(\alpha, \lambda)_{\text{CV}}$ are shown in (main) Figure 2. Although this methodology utilises cross-validation in order to tune the parameters λ and α, the final model is fit on the full data. Stability selection [32] is an alternative approach based on subsampling, which under certain assumptions yields a bound on the expected number of falsely selected variables, $\mathbb{E}[V]$. Based on an assumption of exchangeability, we have

$$\mathbb{E}[V] \leq \frac{1}{2\pi_{\text{thr}} - 1} \frac{q_{\Lambda}^2}{p}$$

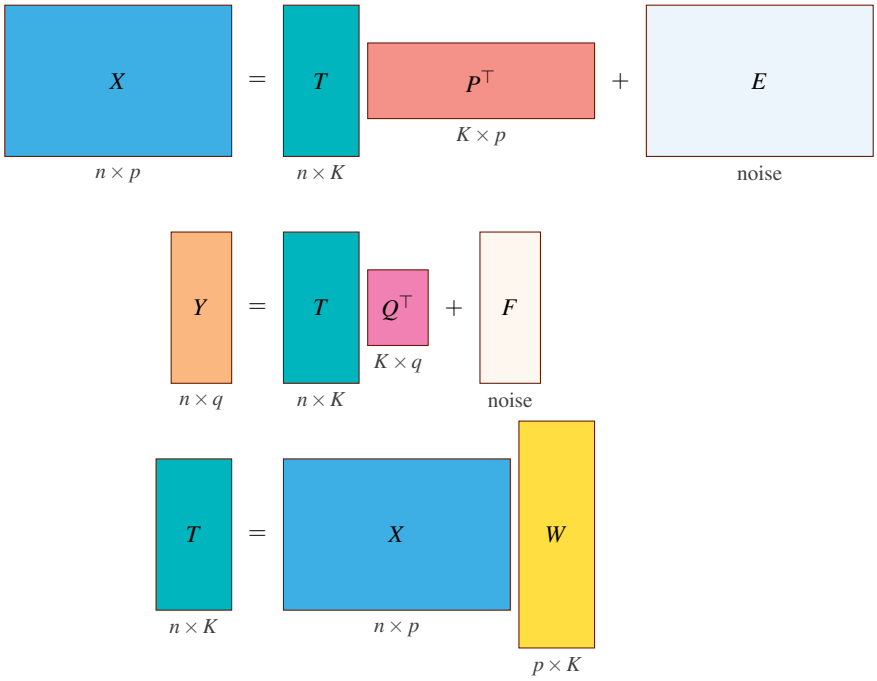
where q_{Λ}^2 is computed from the set Λ of considered regularisation parameters, and π_{thr} can be chosen to bound $\mathbb{E}[V]$ at a desired number. In practice, this bound is not strict, but a variant method called complementary pairs stability selection [33] relaxes some of the model assumptions whilst being less conservative. We therefore use the latter approach, through the implementation in the R package `stabs` [34].

S4.3 Projection to latent structures

We use projection to latent structures (PLS; also called partial least squares) to fit the multivariate multiple regression model $Y = BX + F$, where Y is an $n \times q$ matrix of q response variables (disease status and HDL function), X is an $n \times p$ matrix of p predictor variables (clinical data, lipidomics, proteomics, and NMR spectroscopy), and F is a $n \times q$ matrix of errors. This method is apt for modelling multicollinearity amongst multivariate responses and predictors. The goal is to find a latent K-dimensional decomposition of X and Y through a matrix T, namely

$$X = TP^\top + E$$
$$Y = TQ^\top + F$$

with matrix dimensions illustrated in the following figure.



Suppose that X and Y are centred. Then T, P, and Q are not identifiable, so we are more interested in the column space of T. We have $T = X[w_1, \ldots, w_K]$. Once T is constructed, $Q^\top = (T^\top T)^{-1} T^\top Y$. Recall that $Y = XB + F$, but also $Y = TQ^\top + F$, so $XB = TQ^\top$, meaning that $B = X^{-1} TQ^\top = WQ^\top$. In the SIMPLS formulation [35], for $k = 1, \ldots, K$, we optimise W's kth column

$$w_k = \arg\max_w \{w^\top \operatorname{Cov}(X,Y) \operatorname{Cov}(X,Y)^\top w\} \quad \text{s.t.} \quad w^\top w = 1 \text{ and } w^\top \operatorname{Cov}(X) w_j = 0 \text{ for } j < k.$$

We use sparse PLS [36], which builds on ideas from sparse PCA [37, 38], to perform dimensionality reduction and variable selection simultaneously. This yields an estimate of the K latent components $\hat{T} = X\hat{W}$, which in turn allows us to estimate the loadings $\hat{Q} = \min_Q \|Y - \hat{T}Q^\top\|_2$. Finally, the PLS estimate of the multivariate regression coefficients is

$$\hat{B} = \hat{W}\hat{Q}^\top.$$

We need to select the number K of latent components, as well as the amount of regularisation. This is implemented via cross-validation in the R package `spls` [39]. By performing 1000 bootstrap runs of the sparse PLS regression with parameters tuned using cross-validation, we exclude coefficients whose bootstrap 95% confidence interval contains zero. Since we performed model selection before bootstrapping, the resulting confidence intervals will be compromised, and we must interpret them with caution.

S4.4 Gaussian graphical model

Let $X = (X_1, \ldots, X_p)$ denote the random vector of variables described in the experimental design. Each X_j is a random variable corresponding to a clinical covariate, a functional bioassay, or the abundance of a certain protein or lipid species. For our purposes of finding pairs of variables which may be causally linked, it is clear that we should prioritise candidates by their partial correlation rather than their marginal correlation. The partial correlation is a measure of the linear dependence between a pair of variables when controlling for all other variables in the data, thereby adjusting for the influence of observed confounders. Granted, this procedure will be misleading in case of conditioning on a collider, that is, a common effect X_j of two variables X_i and X_k. In such a case, where the causal graph is $X_i \to X_j \leftarrow X_k$, conditioning on X_j makes X_i and X_k dependent, even though they are not causally linked by a directed path. Nevertheless, we consider conditional dependence relationships as candidates for causal links, keeping in mind that they may indeed be spurious.

Let $\mathscr{G} = (V, E)$ be an undirected graph whose vertices $V = \{1, \ldots, p\}$ are associated with the random variables in X. We want to find the edge set E such that a pair $\{j,k\} \in E$ is connected in $\mathscr{G}$ if and only if X_j and X_k are dependent conditioned on all other variables in X. In other words, we want to find the graph $\mathscr{G}$ such that X forms a Markov random field with respect to $\mathscr{G}$ (through the pairwise Markov property).

Consider a multivariate normal random vector $Z \sim \mathscr{N}(\mu, \Sigma)$ with mean $\mu \in \mathbb{R}^p$ and regular covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$, meaning that the precision matrix $\Theta := \Sigma^{-1}$ is well-defined. It can be shown [40] (proposition 5.2) that $Z_j \perp\!\!\!\perp Z_k \mid Z_{V \setminus \{j,k\}}$ if and only if $\Theta_{jk} = 0$. In other words, any pair (Z_j, Z_k) of random variables in Z is conditionally independent given all other variables in Z if and only if $\Theta_{jk} = 0$. Specifically, the partial correlation between Z_j and Z_k given $Z_{V \setminus \{j,k\}}$ is

$$\rho_{j,k|V\setminus\{j,k\}} = \frac{-\Theta_{jk}}{\sqrt{\Theta_{jj}\Theta_{kk}}}.$$

Thus, to find an edge set E such that Z is a Markov random field with respect to $\mathscr{G}$, we need only invert Σ, and let edges correspond to the nonzero entries of Θ.

Given samples $z^{(1)}, \ldots, z^{(n)} \in \mathbb{R}^p$ from Z, the log likelihood $\ell(\mu, \Sigma) = \text{const.} - \frac{n}{2}\log\det\Sigma - \frac{n}{2}\operatorname{tr}\left\{\Sigma^{-1}S\right\} - \frac{n}{2}(\bar{z} - \mu)^\top\Sigma^{-1}(\bar{z} - \mu)$ where $\bar{z} := \frac{1}{n}\sum_{i=1}^n z^{(i)}$ is the sample mean. Since Σ^{-1} is positive definite, the maximum likelihood estimate (MLE) of μ is $\bar{z}$. Given this, the MLE of Σ is found by solving

$$\underset{\tilde{\Sigma} \in \mathbb{S}^p_{\succ 0}}{\arg\max} \left\{ -\log\det\tilde{\Sigma} - \operatorname{tr}(S\tilde{\Sigma}^{-1}) \right\}$$

where $S := \frac{1}{n}\sum_{i=1}^n (x^{(i)} - \bar{x})(x^{(i)} - \bar{x})^\top$ is the sample covariance matrix, and $\mathbb{S}^p_{\succ 0}$ is the set of all positive definite symmetric matrices in $\mathbb{R}^{p \times p}$. In the high-dimensional case where $p > n$, S has $p - n$ vanishing eigenvalues, and is thus rank-deficient, making the optimisation problem ill-conditioned [41] (section 12.3.2). If, however, the underlying graph is sparse, regularisation can make the problem identifiable [42]. The graphical lasso [43] performs a regularised estimation of the precision matrix $\Theta := \Sigma^{-1}$ by solving

$$\underset{\tilde{\Theta} \in \mathbb{S}^p_{\succ 0}}{\arg\max} \left\{ \log\det\tilde{\Theta} - \operatorname{tr}(S\tilde{\Theta}) - \lambda \sum_{i \neq j}^{p} |\tilde{\Theta}_{ij}| \right\}. \tag{6}$$

The regularisation parameter $\lambda > 0$ controls the sparsity of Θ (the edges of $\mathscr{G}$) and can be tuned using cross-validation. Alternatively, stability selection [32] can be used to bound the expected number of falsely selected edges. In our experience, the former tends to be too liberal, whereas the latter tends to be rather conservative, even in the less stringent complementary-pairs formulation [33]. Furthermore, although stability selection yields a selection probability for each edge in the graph, the actual edge weight is lost in the subsampling. Building on the idea of stability selection, StARS [44] uses subsampling to determine the least amount of regularisation which yields a model that is both sparse and robust to random sampling. This suits our purposes, especially since the StARS-optimal model $\hat{\Theta}$ allows us to estimate the partial correlation coefficients between variables in the model.

We normalised our data and adjusted for batch effects using linear mixed models; see §S4.1.2. However, the continuous variables tend to have marginal distributions with heavier tails than those of a normal distribution. To handle this,

we made use of the *nonparanormal* distribution $\mathrm{NPN}(\mu, \Sigma, f)$. We say that $X \sim \mathrm{NPN}(\mu, \Sigma, f)$ if there exist monotone and differentiable functions $f_1, \ldots, f_p$ such that $\big(f_1(X_1), \ldots, f_p(X_p)\big) \sim \mathcal{N}(\mu, \Sigma)$. In such a case, the nonparanormal distribution is a Gaussian copula, and it can be shown that $X_j \perp\!\!\!\perp X_k \mid X_{V \setminus \{j,k\}}$ if and only if $(\Sigma^{-1})_{jk} = 0$ [45]. For identifiability, we require that the f_j preserve means and variances, that is $\mu_j = \mathbb{E}X_j = \mathbb{E}Z_j$ and $\Sigma_{jj} = \mathrm{Var}\, X_j = \mathrm{Var}\, Z_j$ for all $j \in V$. For the marginal distribution function F_j of X_j, we have

$$F_j(x) = P(X_j \leq x) = P\big(Z_j \leq f_j(x)\big) = \Phi\left(\frac{f_j(x) - \mu_j}{\Sigma_{jj}}\right)$$

where Φ is the cumulative distribution function of the standard normal distribution. Since we centre and scale our data in advance, this means that $f_j(x) = \Phi^{-1}\big(F_j(x)\big)$.

From our samples $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^p$, we estimate F_j by the empirical distribution function of X_j, so that $\hat{F}_j(t) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}[x_j^{(i)} \leq t]$, where 1 has been added to the denominator as a continuity correction. We estimate each f_j by $\hat{f}_j(x) = \Phi^{-1}\big(\hat{F}_j(x)\big)$. Under these transformations, the nonparanormal sample covariance matrix is

$$S(\hat{f}) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}(X^{(i)}) \hat{f}(X^{(i)})^{\top}.$$

By replacing S in the standard graphical lasso estimator (6) with $S(\hat{f})$, we solve for a precision matrix $\hat{\Theta}$ whose corresponding edge set $\hat{E} = \{(j,k) : \hat{\Theta}_{jk} \neq 0\}$ makes our variables X a Markov random field over $\hat{G} = (V, \hat{E})$ [45]. In other words, $X_j \perp\!\!\!\perp X_k \mid X_{V \setminus \{j,k\}}$ if and only if $\hat{\Theta}_{jk} = 0$.

S5 Mixed-effects models

As described in §S4.1.3, we used mixed-effects models to analyse the data from functional bioassays. By including fixed effects for indicator variables of CHD and T2DM, we assessed disease association with each HDL function (considered separately). For a given HDL function, the estimated effect of T2DM, say, reflects the average change in the respective bioassay read-out attributed to the presence of T2DM — other things (plate, CHD status, etc.) being equal. In some assays, the HDL isolation dates were either partially or entirely aliased by the plates, in which case the aliased levels were dropped to make the model identifiable. For the respiration assays (Supplementary Table S13), disease effects were contrasted against those of rotenone/antimycine, in order to show only mitochondrial respiration.

All estimated effects of the mixed models are recorded in Supplementary Tables S5 through S13, with effects on the scale measured by the given bioassay. Through dividing an estimated effect by the residual standard deviation of the corresponding model, we estimate standardised effect sizes. The following Supplementary Table S4 records the standardised effect sizes of diabetes and CHD on HDL function. For each disease effect, the 2.5% and 97.5% percentiles of the bootstrap confidence interval are listed. If a confidence interval does not contain zero, then the effect is significant at the 5% level, indicated by an asterisk.

assay	Effect	Estimate	2.5 %	97.5 %	Significant
apoptosis HAEC	HDL T2DM	0.498	0.15	0.846	*
apoptosis HAEC	HDL CHD	−0.438	−0.953	0.0763	
apoptosis INS1e	HDL T2DM	−0.149	−0.438	0.14	
apoptosis INS1e	HDL CHD	0.309	−0.129	0.748	
CEC apoB-free plasma	HDL T2DM	0.0521	−0.879	0.983	
CEC apoB-free plasma	HDL CHD	−0.6	−1.83	0.634	
CEC HDL	HDL T2DM	−0.18	−0.629	0.268	
CEC HDL	HDL CHD	−0.323	−0.977	0.331	
10min-MMP C2C12	HDL T2DM	0.134	−0.121	0.389	
10min-MMP C2C12	HDL CHD	−0.101	−0.479	0.277	
1hr-MMP C2C12	HDL T2DM	0.172	−0.044	0.388	
1hr-MMP C2C12	HDL CHD	−0.0566	−0.385	0.271	
2hr-MMP C2C12	HDL T2DM	0.189	−0.0231	0.402	
2hr-MMP C2C12	HDL CHD	−0.00747	−0.334	0.319	
10min CCCP-MMP C2C12	HDL T2DM	0.283	0.00683	0.559	*
10min CCCP-MMP C2C12	HDL CHD	−0.0419	−0.44	0.356	
respiration basal	HDL T2DM	−0.0369	−0.399	0.325	
respiration basal	HDL CHD	−0.057	−0.626	0.512	
respiration cAMP	HDL T2DM	−0.0656	−0.492	0.361	
respiration cAMP	HDL CHD	0.0527	−0.516	0.622	
respiration max	HDL T2DM	−0.394	−0.768	−0.0192	*
respiration max	HDL CHD	0.677	0.108	1.25	*
respiration uncoupled	HDL T2DM	−0.0133	−0.38	0.354	
respiration uncoupled	HDL CHD	−0.0357	−0.638	0.567	

Supplementary Table S4: Standardised effects sizes of diseased HDL on the assayed cellular functions.

Effect	Estimate	2.5 %	97.5 %	Significant
sd_(Intercept)|HDL No	6.9	5.56	8.24	*
sigma	8.76	8.04	9.49	*
(Intercept)	55.5	48.9	62.1	*
plateplate 2	−10.3	−15.5	−5.03	*
plateplate 3	12.3	7.08	17.4	*
plateplate 4	−2.24	−7.11	2.62	
plateplate 5	−9.84	−15.2	−4.51	*
plateplate 6	−8.28	−13.4	−3.16	*
plateplate 7	−22.7	−28.2	−17.2	*
Diabetes	4.37	1.32	7.42	*
CHD	−3.84	−8.35	0.669	
hospitalZurich	−1.8	−7.31	3.71	
Female	−1.92	−4.93	1.08	
rotor 2	2.67	−0.0851	5.42	

Supplementary Table S5: Mixed-effects model estimates with **apoptosis HAEC** as response.

Effect	Estimate	2.5 %	97.5 %	Significant
sd_(Intercept)|HDL No	6.95	5.15	8.76	*
sigma	11.9	10.9	12.8	*
(Intercept)	79	71.5	86.5	*
plateplate 2	6.98	0.919	13	*
plateplate 3	−7.66	−13.7	−1.64	*
plateplate 4	3.78	−1.88	9.44	
plateplate 5	−13	−19	−6.95	*
plateplate 6	−22.4	−28.4	−16.5	*
plateplate 7	−20.4	−26.6	−14.2	*
Diabetes	−1.77	−5.19	1.66	
CHD	3.66	−1.54	8.86	
hospitalZurich	0.664	−5.93	7.26	
Female	−1.3	−4.69	2.08	
rotor 2	−3.41	−6.58	−0.231	*

Supplementary Table S6: Mixed-effects model estimates with **apoptosis INS1e** as response.

Effect	Estimate	2.5 %	97.5 %	Significant
sd_(Intercept)|HDL No	14.1	12.4	15.8	*
sigma	6.23	5.71	6.75	*
(Intercept)	91.3	80.3	102	*
plateplate 10	−21.9	−30.6	−13.1	*
plateplate 11	−9.46	−18.2	−0.76	*
plateplate 2	−6.53	−27.2	14.2	
plateplate 3	−3.72	−30	22.5	
plateplate 4	12.5	−17.8	42.7	
plateplate 9	−24.3	−43.2	−5.34	*
Diabetes	0.325	−5.47	6.12	
CHD	−3.74	−11.4	3.95	
hospitalZurich	1.53	−7.7	10.8	
Female	2.72	−2.74	8.18	
rotor 2	−3.46	−8.68	1.76	
‘Start isolation‘20.9.	14.3	−3.28	31.9	
‘Start isolation‘23.4.	−1.95	−23.9	20	
‘Start isolation‘26.4.	−5.27	−32.5	22	
‘Start isolation‘30.4.	−21	−51.6	9.61	

Supplementary Table S7: Mixed-effects model estimates with **CEC apoB-free plasma** as response.

Effect	Estimate	2.5 %	97.5 %	Significant
sd_(Intercept)|HDL No	6.43	5.42	7.44	*
sigma	6.66	6.12	7.19	*
(Intercept)	49.5	36.7	62.3	*
plateplate 13	3.38	−7.56	14.3	
plateplate 14	−0.188	−14	13.6	
plateplate 5	23.3	10.2	36.4	*
plateplate 6	−3.15	−23.7	17.4	
plateplate 7	−14.8	−37.7	7.98	
plateplate 8	22.7	−1.23	46.6	
Diabetes	−1.2	−4.19	1.79	
CHD	−2.15	−6.5	2.2	
hospitalZurich	−1.68	−6.32	2.97	
Female	3.4	0.588	6.21	*
rotor 2	−2.99	−5.49	−0.496	*
‘Start isolation‘20.9.	4.6	−8.22	17.4	
‘Start isolation‘23.4.	29.6	13.9	45.2	*
‘Start isolation‘24.9.	−3.88	−11	3.27	
‘Start isolation‘26.4.	36	17.2	54.8	*
‘Start isolation‘30.4.	19.7	−0.391	39.9	

Supplementary Table S8: Mixed-effects model estimates with **CEC HDL** as response.

Effect	Estimate	2.5 %	97.5 %	Significant
sd_(Intercept)|HDL No	2.75	1.88	3.63	*
sigma	6.37	5.96	6.79	*
(Intercept)	−19.6	−23.1	−16.1	*
plateplate 10	−4.59	−7.85	−1.34	*
plateplate 11	−11.5	−15.1	−7.85	*
plateplate 12	3.39	−0.143	6.93	
plateplate 13	−0.0156	−3.41	3.38	
plateplate 14	−6.13	−9.8	−2.47	*
plateplate 2	3.18	0.803	5.55	*
plateplate 3	4.07	1.06	7.07	*
plateplate 4	−2.54	−5.6	0.526	
plateplate 5	8.17	5.04	11.3	*
plateplate 6	−2.08	−5.17	1.02	
plateplate 7	2.06	−1.01	5.12	
plateplate 8	3.26	0.304	6.22	*
plateplate 9	−5.99	−9.47	−2.51	*
Diabetes	0.854	−0.772	2.48	
CHD	−0.644	−3.05	1.76	
hospitalZurich	−0.134	−2.89	2.62	
Female	−0.0934	−1.72	1.53	
rotor 2	1.1	−0.777	2.97	

Supplementary Table S9: Mixed-effects model estimates with **10min-MMP C2C12** as response.

Effect	Estimate	2.5 %	97.5 %	Significant
sd_(Intercept)|HDL No	1.6	0.00221	3.2	*
sigma	7.68	7.18	8.17	*
(Intercept)	−32.6	−36.5	−28.8	*
plateplate 10	−3.25	−6.6	0.0963	
plateplate 11	0.742	−3.14	4.63	
plateplate 12	7.82	4.09	11.5	*
plateplate 13	4.93	1.25	8.6	*
plateplate 14	0.436	−3.3	4.17	
plateplate 2	4.77	1.91	7.63	*
plateplate 3	0.538	−2.82	3.89	
plateplate 4	0.552	−2.72	3.82	
plateplate 5	5.96	2.58	9.34	*
plateplate 6	−1.61	−5.16	1.95	
plateplate 7	1.02	−2.31	4.34	
plateplate 8	2.3	−0.974	5.58	
plateplate 9	−4.29	−7.87	−0.716	*
Diabetes	1.32	−0.338	2.98	
CHD	−0.435	−2.95	2.08	
hospitalZurich	0.501	−2.37	3.37	
Female	−0.55	−2.22	1.12	
rotor 2	1.27	−0.651	3.18	

Supplementary Table S10: Mixed-effects model estimates with **1hr-MMP C2C12** as response.

Effect	Estimate	2.5 %	97.5 %	Significant
sd_(Intercept)|HDL No	1.63	2.01×10^{-7}	3.26	*
sigma	8.05	7.53	8.56	*
(Intercept)	−36.3	−40.3	−32.3	*
plateplate 10	−1.23	−4.71	2.25	
plateplate 11	−1.7	−5.71	2.31	
plateplate 12	5.72	1.87	9.58	*
plateplate 13	6.04	2.22	9.87	*
plateplate 14	−0.0687	−3.93	3.79	
plateplate 2	3.52	0.526	6.52	*
plateplate 3	−2.1	−5.61	1.4	
plateplate 4	2.27	−1.11	5.64	
plateplate 5	1.96	−1.57	5.49	
plateplate 6	−1.72	−5.42	1.97	
plateplate 7	−1.3	−4.8	2.2	
plateplate 8	−0.416	−3.83	3	
plateplate 9	−3.41	−7.1	0.276	
Diabetes	1.52	−0.186	3.23	
CHD	−0.0601	−2.69	2.57	
hospitalZurich	0.939	−2.02	3.9	
Female	−0.889	−2.63	0.849	
rotor 2	0.968	−1.02	2.96	

Supplementary Table S11: Mixed-effects model estimates with **2hr-MMP C2C12** as response.

Effect	Estimate	2.5 %	97.5 %	Significant
sd_(Intercept)|HDL No	2.54	1.9	3.19	*
sigma	4.85	4.53	5.16	*
(Intercept)	−69.9	−72.8	−67	*
plateplate 10	15.7	13.1	18.4	*
plateplate 11	22	19.1	25	*
plateplate 12	31.2	28.4	34	*
plateplate 13	11.2	8.38	13.9	*
plateplate 14	9.61	6.66	12.6	*
plateplate 2	−0.193	−2	1.61	
plateplate 3	9.59	7.15	12	*
plateplate 4	9.15	6.6	11.7	*
plateplate 5	9.2	6.57	11.8	*
plateplate 6	9.01	6.59	11.4	*
plateplate 7	7.23	4.77	9.68	*
plateplate 8	9.17	6.83	11.5	*
plateplate 9	14.2	11.3	17.2	*
Diabetes	1.37	0.0331	2.71	*
CHD	−0.203	−2.13	1.73	
hospitalZurich	−0.0513	−2.29	2.19	
Female	−0.376	−1.71	0.96	
rotor 2	−0.569	−2.1	0.957	

Supplementary Table S12: Mixed-effects model estimates with **10min CCCP-MMP C2C12** as response.

Effect	Estimate	2.5 %	97.5 %	Significant
sd_(Intercept)|key:'HDL No'	7.36	6.88	7.84	*
sigma	10.8	10.7	11	*
(Intercept)	45.5	38.7	52.3	*
basal	56.4	46	66.9	*
cAMP	76.1	66.5	85.7	*
max	155	145	164	*
uncoupled	15.3	4.91	25.7	*
Diabetes	−0.359	−3.33	2.61	
CHD	0.75	−3.84	5.34	
measurementMeasurement 10	16.4	15.3	17.5	*
measurementMeasurement 11	8.43	7.3	9.56	*
measurementMeasurement 12	3.02	1.96	4.09	*
measurementMeasurement 14	−2.84	−3.99	−1.69	*
measurementMeasurement 15	0.847	−0.345	2.04	
measurementMeasurement 2	−8.6	−9.69	−7.5	*
measurementMeasurement 3	−11.2	−12.4	−10.1	*
measurementMeasurement 4	3.85	2.7	4.99	*
measurementMeasurement 5	1.91	0.757	3.05	*
basal:Diabetes	−0.4	−4.32	3.52	
cAMP:Diabetes	−0.711	−5.33	3.91	
max:Diabetes	−4.27	−8.33	−0.209	*
uncoupled:Diabetes	−0.144	−4.12	3.83	
basal:CHD	−0.617	−6.78	5.55	
cAMP:CHD	0.572	−5.6	6.74	
max:CHD	7.33	1.17	13.5	*
uncoupled:CHD	−0.387	−6.92	6.15	
Rotenone/Antimycine:platePlate 10	18.2	−0.151	36.6	
basal:platePlate 10	16.9	0.0462	33.8	*
cAMP:platePlate 10	6.28	−11.9	24.5	
max:platePlate 10	34.7	15.9	53.6	*
uncoupled:platePlate 10	14.3	−2.75	31.3	
Rotenone/Antimycine:platePlate 11	7.9	0.597	15.2	*
basal:platePlate 11	−0.195	−7.35	6.96	
cAMP:platePlate 11	−3.13	−10.2	3.96	
max:platePlate 11	−5.97	−12.5	0.544	
uncoupled:platePlate 11	0.412	−6.51	7.33	
Rotenone/Antimycine:platePlate 12	20.2	2.43	38	*
basal:platePlate 12	11.9	−5.62	29.4	
cAMP:platePlate 12	2.76	−16.8	22.3	
max:platePlate 12	34.2	16.8	51.7	*
uncoupled:platePlate 12	14.3	−3.33	31.9	
Rotenone/Antimycine:platePlate 13	10.1	3.03	17.2	*
basal:platePlate 13	17.1	9.92	24.2	*
cAMP:platePlate 13	1.55	−5.52	8.61	
max:platePlate 13	−4.81	−11.4	1.79	
uncoupled:platePlate 13	12	5.42	18.5	*
Rotenone/Antimycine:platePlate 14	19.1	0.576	37.6	*
basal:platePlate 14	19.5	2.09	36.8	*
cAMP:platePlate 14	6.01	−13.3	25.3	
max:platePlate 14	1.07	−17.6	19.7	
uncoupled:platePlate 14	21.7	4.3	39.1	*

Rotenone/Antimycine:platePlate 2	6.57	−10.4	23.6	
basal:platePlate 2	0.246	−16.5	17	
cAMP:platePlate 2	15.9	−2.57	34.3	
max:platePlate 2	29.3	11.7	46.9	*
uncoupled:platePlate 2	4.97	−11.3	21.2	
Rotenone/Antimycine:platePlate 3	4.32	−2.87	11.5	
basal:platePlate 3	−3.6	−10.2	2.96	
cAMP:platePlate 3	14	7.42	20.6	*
max:platePlate 3	4.48	−2.3	11.3	
uncoupled:platePlate 3	0.134	−6.61	6.88	
Rotenone/Antimycine:platePlate 4	5.05	−13.5	23.6	
basal:platePlate 4	−3.5	−21.4	14.4	
cAMP:platePlate 4	13.2	−5.98	32.4	
max:platePlate 4	2.44	−16	20.9	
uncoupled:platePlate 4	2.76	−14.8	20.3	
Rotenone/Antimycine:platePlate 5	5.5	−1.56	12.6	
basal:platePlate 5	5.18	−1.72	12.1	
cAMP:platePlate 5	12.4	5.47	19.4	*
max:platePlate 5	77.3	70	84.6	*
uncoupled:platePlate 5	8.25	1.54	15	*
Rotenone/Antimycine:platePlate 6	9.11	−8.62	26.8	
basal:platePlate 6	7.75	−10.4	25.9	
cAMP:platePlate 6	18.1	−1.37	37.5	
max:platePlate 6	92.7	74.1	111	*
uncoupled:platePlate 6	13.5	−4.18	31.2	
Rotenone/Antimycine:platePlate 7	6.11	−1.02	13.2	
basal:platePlate 7	−2.21	−8.75	4.34	
cAMP:platePlate 7	22.9	15.9	29.8	*
max:platePlate 7	36	28.7	43.3	*
uncoupled:platePlate 7	4.67	−2.04	11.4	
Rotenone/Antimycine:platePlate 8	6.37	−11.7	24.4	
basal:platePlate 8	−7.21	−24.5	10.1	
cAMP:platePlate 8	13.4	−4.85	31.7	
max:platePlate 8	12.4	−5.14	30	
uncoupled:platePlate 8	3.01	−14.7	20.8	
Rotenone/Antimycine:platePlate 9	11.3	4.29	18.2	*
basal:platePlate 9	12	5.63	18.5	*
cAMP:platePlate 9	−9.66	−16.5	−2.85	*
max:platePlate 9	28.4	21.7	35.2	*
uncoupled:platePlate 9	6.84	0.195	13.5	*
Rotenone/Antimycine:hospitalZurich	0.0723	−5.32	5.46	
basal:hospitalZurich	−1.48	−6.52	3.56	
cAMP:hospitalZurich	−1.22	−6.86	4.42	
max:hospitalZurich	3.72	−1.39	8.84	
uncoupled:hospitalZurich	−0.274	−5.68	5.13	
Rotenone/Antimycine:Female	0.7	−2.23	3.63	
basal:Female	1.43	−1.41	4.27	
cAMP:Female	1.52	−1.32	4.35	
max:Female	3.33	0.465	6.2	*
uncoupled:Female	0.853	−2	3.71	
Rotenone/Antimycine:rotor 2	−1.51	−18.6	15.6	
basal:rotor 2	0.947	−14.8	16.7	
cAMP:rotor 2	−2.14	−19.5	15.3	

max:rotor 2	−11.5	−27.7	4.73
uncoupled:rotor 2	−2.09	−18.2	14

Supplementary Table S13: Mixed-effects model estimates with **adipocyte respiration** as response. The effect of each type of respiration (basal, uncoupled, cAMP-stimulated, and maximal) is contrasted against that of rotenone/antimycine (absorbed into the model intercept), in order to show only mitochondrial respiration.

References

[1] Philip Garner and Jung Min Park. 1,1,-Dimethylethyl (S)- or (R)-4-Formyl-2,2-Dimethyl-3-Oxazolidinecarboxylate: A Useful Serinal Derivative. *Org. Synth.*, 70:18, 1992.

[2] Peter Herold. Synthesis of D-Erythro- and D-Threo-Sphingosine Derivatives From L-Serine. *Helvetica Chimica Acta*, 71(2):354–362, 1988.

[3] Hans Rudolf Pfaendler and Volker Weimar. Synthesis of racemic ethanolamine plasmalogen. *Synthesis*, 1996(11):1345–1349, 1996.

[4] Hagan Bayley, David N. Standring, and Jeremy R. Knowles. Propane-1,3-dithiol: A selective reagent for the efficient reduction of alkyl and aryl azides to amines. *Tetrahedron Letters*, 19(39):3633–3634, January 1978.

[5] Kenji Mori, Hidenori Watanabe, Mitsuhiko Fujiwhara, and Shigefumi Kuwahara. Pheromone synthesis, CXXII. (E)- and (Z)-tetradecenyl formate, potent sex pheromone mimics against the yellow peach moth. *Liebigs Annalen der Chemie*, 1990(12):1257–1259, 1990.

[6] Ludovic C. Gillet, Pedro Navarro, Stephen Tate, Hannes Röst, Nathalie Selevsek, Lukas Reiter, Ron Bonner, and Ruedi Aebersold. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell Proteomics*, 11(6):O111.016717, June 2012.

[7] J. K. Eng, A. L. McCormack, and J. R. Yates. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.*, 5(11):976–989, November 1994.

[8] Viktoria Dorfer, Peter Pichler, Thomas Stranzl, Johannes Stadlmann, Thomas Taus, Stephan Winkler, and Karl Mechtler. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J. Proteome Res.*, 13(8):3679–3684, August 2014.

[9] Roland Bruderer, Oliver M. Bernhardt, Tejas Gandhi, Saša M. Miladinović, Lin-Yang Cheng, Simon Messner, Tobias Ehrenberger, Vito Zanotelli, Yulia Butscheid, Claudia Escher, Olga Vitek, Oliver Rinner, and Lukas Reiter. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell Proteomics*, 14(5):1400–1410, May 2015.

[10] Gerhard Liebisch, Bernd Lieser, Jan Rathenberg, Wolfgang Drobnik, and Gerd Schmitz. High-throughput quantification of phosphatidylcholine and sphingomyelin by electrospray ionization tandem mass spectrometry coupled with isotope correction algorithm. *Biochim. Biophys. Acta*, 1686(1-2):108–117, November 2004.

[11] Gerhard Liebisch, Wolfgang Drobnik, Bernd Lieser, and Gerd Schmitz. High-throughput quantification of lysophosphatidylcholine by electrospray ionization tandem mass spectrometry. *Clin. Chem.*, 48(12):2217–2224, December 2002.

[12] B. Brügger, G. Erben, R. Sandhoff, F. T. Wieland, and W. D. Lehmann. Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.*, 94(6):2339–2344, March 1997.

[13] Vitali Matyash, Gerhard Liebisch, Teymuras V. Kurzchalia, Andrej Shevchenko, and Dominik Schwudke. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J. Lipid Res.*, 49(5):1137–1146, May 2008.

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

[15] G. Liebisch, W. Drobnik, M. Reil, B. Trümbach, R. Arnecke, B. Olgemöller, A. Roscher, and G. Schmitz. Quantitative measurement of different ceramide species from crude cellular extracts by electrospray ionization tandem mass spectrometry (ESI-MS/MS). *J. Lipid Res.*, 40(8):1539–1546, August 1999.

[16] Gerhard Liebisch, Marion Binder, Rainer Schifferer, Thomas Langmann, Berta Schulz, and Gerd Schmitz. High throughput quantification of cholesterol and cholesteryl ester by electrospray ionization tandem mass spectrometry (ESI-MS/MS). *Biochim. Biophys. Acta*, 1761(1):121–128, January 2006.

[17] Max Scherer, Gerd Schmitz, and Gerhard Liebisch. Simultaneous quantification of cardiolipin, bis(monoacylglycero)phosphate and their precursors by hydrophilic interaction LC-MS/MS including correction of isotopic overlap. *Anal. Chem.*, 82(21):8794–8799, November 2010.

[18] Gerhard Liebisch, Juan Antonio Vizcaíno, Harald Köfeler, Martin Trötzmüller, William J. Griffiths, Gerd Schmitz, Friedrich Spener, and Michael J. O. Wakelam. Shorthand notation for lipid structures derived from mass spectrometry. *J. Lipid Res.*, 54(6):1523–1530, June 2013.

[19] C Croghan and PP Egeghy. Methods of dealing with values below the limit of detection using SAS. *Southern SAS User Group*, 2003.

[20] R Core Team. R: A Language and Environment for Statistical Computing, 2018.

[21] Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3):1–67, 2011.

[22] Bart J. A. Mertens. Transformation, Normalization, and Batch Effect in the Analysis of Mass Spectrometry Data for Omics Studies. In Susmita Datta and Bart J. A. Mertens, editors, *Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry*, Frontiers in Probability and the Statistical Sciences, pages 1–21. Springer International Publishing, Cham, 2017.

[23] Carl Murie, Brian Sandri, Ann-Sofi Sandberg, Timothy J. Griffin, Janne Lehtiö, Christine Wendt, and Ola Larsson. Normalization of mass spectrometry data (NOMAD). *Advances in Biological Regulation*, 67:128–133, January 2018.

[24] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[25] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics. Springer-Verlag, New York, 2 edition, 2009.

[26] Trevor Hastie. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman and Hall/CRC, New York, 2015.

[27] Xiaolin Yang, Seyoung Kim, and Eric P. Xing. Heterogeneous multitask learning with joint sparsity constraints. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 2151–2159. Curran Associates, Inc., 2009.

[28] Robert Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[29] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

[30] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1):1–22, 2010.

[31] Microsoft and Hong Ooi. glmnetUtils: Utilities for 'Glmnet', 2017.

[32] Nicolai Meinshausen and Peter Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.

[33] Rajen D. Shah and Richard J. Samworth. Variable selection with error control: another look at stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):55–80, 2012.

[34] Benjamin Hofner and Torsten Hothorn. stabs: Stability Selection with Error Control, 2017.

[35] Sijmen de Jong. SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):251–263, March 1993.

[36] Hyonho Chun and Sündüz Keleş. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):3–25, 2010.

[37] Ian T. Jolliffe, Nickolay T. Trendafilov, and Mudassir Uddin. A Modified Principal Component Technique Based on the LASSO. *Journal of Computational and Graphical Statistics*, 12(3):531–547, September 2003.

[38] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, June 2006.

[39] Dongjun Chung, Hyonho Chun, and Sunduz Keles. spls: Sparse Partial Least Squares (SPLS) Regression and Classification, 2019.

[40] Steffen L Lauritzen. *Graphical Models*. Number 17 in Oxford Statistical Science Series. Oxford University Press, 1996.

[41] Marloes Maathuis, Mathias Drton, Steffen Lauritzen, and Martin Wainwright, editors. *Handbook of Graphical Models*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2018.

[42] Adam J. Rothman, Peter J. Bickel, Elizaveta Levina, and Ji Zhu. Sparse permutation invariant covariance estimation. *Electron. J. Statist.*, 2:494–515, 2008.

[43] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, July 2008.

[44] Han Liu, Kathryn Roeder, and Larry Wasserman. Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 1432–1440. Curran Associates, Inc., 2010.

[45] Han Liu, John Lafferty, and Larry Wasserman. The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs. *Journal of Machine Learning Research*, 10(Oct):2295–2328, 2009.