SUPPORTING INFORMATION

An unbiased lipid phenotyping approach to study the genetic determinants of lipids and their association with coronary heart disease risk factors

Eric L. Harshfield[1,2†], Albert Koulman[3†], Daniel Ziemek[4], Luke Marney[5], Eric B. Fauman[6], Dirk S. Paul[1], David Stacey[1], Asif Rasheed[7], Jung-Jin Lee[14], Nabi Shah[7,8], Sehrish Jabeen[7], Atif Imran[7], Shahid Abbas[9], Zoubia Hina[9], Nadeem Qamar[10], Nadeem Hayyat Mallick[11], Zia Yaqoob[12], Tahir Saghir[10], Syed Nadeem Hasan Rizvi[10], Anis Memon[10], Syed Zahed Rasheed[12], Fazal-ur-Rehman Memon[13], Irshad Hussain Qureshi[14], Muhammad Ishaq[12], Philippe Frossard[7], John Danesh[1], Danish Saleheen[7,15], Adam S. Butterworth[1], Angela M. Wood[1‡], and Julian L. Griffin[16*‡]

AUTHOR AFFILIATIONS

1. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, UK
2. Stroke Research Group, Department of Clinical Neurosciences, University of Cambridge, UK
3. Core Metabolomics and Lipidomics Laboratory, National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge, UK
4. Inflammation and Immunology, Pfizer Worldwide Research & Development, Berlin, Germany
5. College of Science and Engineering, Seattle University, Seattle, Washington, USA
6. Genomic Sciences and Technologies, Pfizer Worldwide Research & Development, Cambridge, Massachusetts, USA
7. Center for Non-Communicable Diseases, Karachi, Pakistan
8. Department of Pharmacy, COMSATS Institute of Information Technology, Abbottabad, Pakistan
9. Department of Cardiology, Faisalabad Institute of Cardiology, Faisalabad, Pakistan
10. National Institute of Cardiovascular Disorders, Karachi, Pakistan
11. Punjab Institute of Cardiology, Lahore, Pakistan
12. Karachi Institute of Heart Diseases, Karachi, Pakistan
13. Red Crescent Institute of Cardiology, Hyderabad, Pakistan
14. Department of Medicine, Mayo Hospital, Lahore, Pakistan
15. Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA
16. Department of Biochemistry and Cambridge Systems Biology Centre, University of Cambridge, UK

†These authors contributed equally to this work. ‡These authors jointly directed this work.

***Corresponding Author:** Julian L. Griffin, Department of Biochemistry, Tennis Court Road, University of Cambridge, Cambridge, CB2 1GA, UK. Email: jlg40@cam.ac.uk. Tel: +44 1223 764 922. Fax: +44 1223 333 345.

TABLE OF CONTENTS

Supplementary Methods. Technical discussion on annotation of lipid signals

Lipid signals were annotated to a molecular formula on the basis of the accurate mass. The resolution of 65,000 at 400 *m/z,* as used in this study, allowed for the baseline separation of, for instance, molecular formulae $C_{41}H_{78}NPO_8+H^+$ (*m/z* 744.554) and $C_{42}H_{82}NPO_7+H^+$ (*m/z* 744.590), but was unable to determine if the former was PC(33:2) or PE(36:2) as the species are isobaric. The average mass accuracy error in the measurement of the *m/z* was 0.85 ppm across all intact lipids, with the highest difference for detected lipids of 2.69 ppm for PC ae (37:4). However, this did not mean that only one lipid species contributed to a specific mass-to-charge ratio. For example, the ion that we identified as TG(52:2) with *m/z* 876.802 could be many different triglyceride lipids [e.g. TG(16:0/18:2/18:0), TG(14:0/16:0/22:2) or TG(16:0/18:1/18:1)], all of which have the same molecular weight. Interpretation of species at the chemical formula level allowed us to model changes within lipid pools according to biological processes such as chain elongation and desaturation. We assumed that a given signal peak was likely to be a combination of several lipid species. Our annotation was further based on fragmentation data for the most predominant ion through additional LC-MS/MS analyses.

Because the "open-profiling" approach did not predetermine which lipid species would be detected, it provided data on all ionisable molecules and was therefore very sensitive to contamination, especially of compounds with high ionization efficiencies. In all analyses, we found adipates (*m/z* 371.316) and organophosphates, such as Tris(ditert-butylphenyl)phosphite (*m/z* 647.459) and its oxidation products (*m/z* 663.454), that leached from plastics into the organic solvents. However, using glass-coated well-plates minimized the contact time of the samples with the plastics, and by using blanks and QCs at three different concentrations, we were able to exclude the contaminating ions (approximately half of all the signals) from the final data set. The use of glass-coated well plates was essential to obtain both precise and reliable data. Furthermore, as the method relied on nanoflow, contaminants had minimal impact on the ionization efficiency.

The developed peak-picking algorithm enabled us to process the almost seven thousand data files using parallelization with a processing time of about four minutes per sample per core. The analysis time was greatly sped up by processing each file independently since there was no requirement to load all files jointly into memory to perform the alignment; the ability to perform analyses in parallel also greatly sped up the analysis time. This approach is only suitable to compare similar samples where the same lipids are expected, as it requires known lipids with their target *m/z*. The results required manual curation as in certain cases the target *m/z* was too close to an isotope or adduct of another lipid. We therefore confirmed the identity of all the ions that passed the QC filter, and selected samples were analysed by high-resolution LC-MS/MS to confirm lipid annotations.

Supplementary Table. Summary of genetic associations of principal components of lipids

Principal component	rsID	Chr:Pos (GRCh37)	EA	NEA	Beta	SE	*P*-value	Locus
Second	rs662799	chr11:116663707	A	G	-0.996	0.182	4.30E-08	*APOA5-APOC3*
Third	rs651821	chr11:116662579	T	C	1.2532	0.165	2.70E-14	*APOA5-APOC3*
Third	rs662799	chr11:116663707	A	G	1.2561	0.165	3.00E-14	*APOA5-APOC3*
Third	rs6589566	chr11:116652423	A	G	1.16	0.161	6.60E-13	*APOA5-APOC3*
Third	rs10790162	chr11:116639104	A	G	-1.152	0.161	8.90E-13	*APOA5-APOC3*
Third	rs6589565	chr11:116640237	A	G	-1.149	0.161	1.00E-12	*APOA5-APOC3*
Third	rs964184	chr11:116648917	C	G	1.0656	0.15	1.10E-12	*APOA5-APOC3*
Third	rs2160669	chr11:116647607	T	C	1.1438	0.161	1.10E-12	*APOA5-APOC3*
Third	rs10750096	chr11:116656788	A	C	1.1443	0.162	1.50E-12	*APOA5-APOC3*
Third	rs2075290	chr11:116653296	T	C	1.1359	0.161	1.60E-12	*APOA5-APOC3*
Third	rs3825041	chr11:116631707	T	C	-1.126	0.16	1.60E-12	*APOA5-APOC3*
Third	rs6589564	chr11:116624153	C	G	-1.121	0.16	2.20E-12	*APOA5-APOC3*
Third	rs7930786	chr11:116624727	C	G	-1.114	0.159	2.60E-12	*APOA5-APOC3*
Third	rs1558860	chr11:116607368	A	C	-1.116	0.161	4.10E-12	*APOA5-APOC3*
Third	rs9326246	chr11:116611733	C	G	-1.105	0.16	5.80E-12	*APOA5-APOC3*
Third	rs2072560	chr11:116661826	T	C	-1.151	0.168	7.10E-12	*APOA5-APOC3*
Third	rs2266788	chr11:116660686	A	G	1.0834	0.158	8.00E-12	*APOA5-APOC3*
Third	rs1558861	chr11:116607437	T	C	1.0867	0.16	1.30E-11	*APOA5-APOC3*
Third	rs7483863	chr11:116652491	A	G	-1.065	0.159	2.10E-11	*APOA5-APOC3*
Third	rs11604424	chr11:116651115	T	C	0.8792	0.132	2.60E-11	*APOA5-APOC3*
Third	rs3741298	chr11:116657561	T	C	0.8822	0.134	4.10E-11	*APOA5-APOC3*
Third	rs7350481	chr11:116586283	T	C	-0.897	0.148	1.30E-09	*APOA5-APOC3*
Third	rs180327	chr11:116623659	T	C	0.7048	0.127	2.80E-08	*APOA5-APOC3*
Fourth	rs964184	chr11:116648917	C	G	1.2218	0.143	1.20E-17	*APOA5-APOC3*
Fourth	rs662799	chr11:116663707	A	G	1.2808	0.158	4.70E-16	*APOA5-APOC3*
Fourth	rs3741298	chr11:116657561	T	C	1.0166	0.127	1.50E-15	*APOA5-APOC3*
Fourth	rs651821	chr11:116662579	T	C	1.2499	0.157	1.80E-15	*APOA5-APOC3*
Fourth	rs2072560	chr11:116661826	T	C	-1.234	0.16	1.30E-14	*APOA5-APOC3*
Fourth	rs6589566	chr11:116652423	A	G	1.1734	0.154	2.60E-14	*APOA5-APOC3*
Fourth	rs11604424	chr11:116651115	T	C	0.9542	0.126	3.30E-14	*APOA5-APOC3*
Fourth	rs2160669	chr11:116647607	T	C	1.1589	0.153	4.20E-14	*APOA5-APOC3*
Fourth	rs6589565	chr11:116640237	A	G	-1.159	0.154	4.80E-14	*APOA5-APOC3*
Fourth	rs2075290	chr11:116653296	T	C	1.1567	0.154	5.00E-14	*APOA5-APOC3*
Fourth	rs10790162	chr11:116639104	A	G	-1.159	0.154	5.00E-14	*APOA5-APOC3*
Fourth	rs10750096	chr11:116656788	A	C	1.1604	0.154	5.80E-14	*APOA5-APOC3*
Fourth	rs6589564	chr11:116624153	C	G	-1.129	0.152	1.30E-13	*APOA5-APOC3*
Fourth	rs9326246	chr11:116611733	C	G	-1.134	0.153	1.30E-13	*APOA5-APOC3*
Fourth	rs3825041	chr11:116631707	T	C	-1.126	0.152	1.40E-13	*APOA5-APOC3*
Fourth	rs7930786	chr11:116624727	C	G	-1.123	0.152	1.50E-13	*APOA5-APOC3*
Fourth	rs1558860	chr11:116607368	A	C	-1.126	0.154	2.30E-13	*APOA5-APOC3*
Fourth	rs2266788	chr11:116660686	A	G	1.1044	0.151	3.00E-13	*APOA5-APOC3*

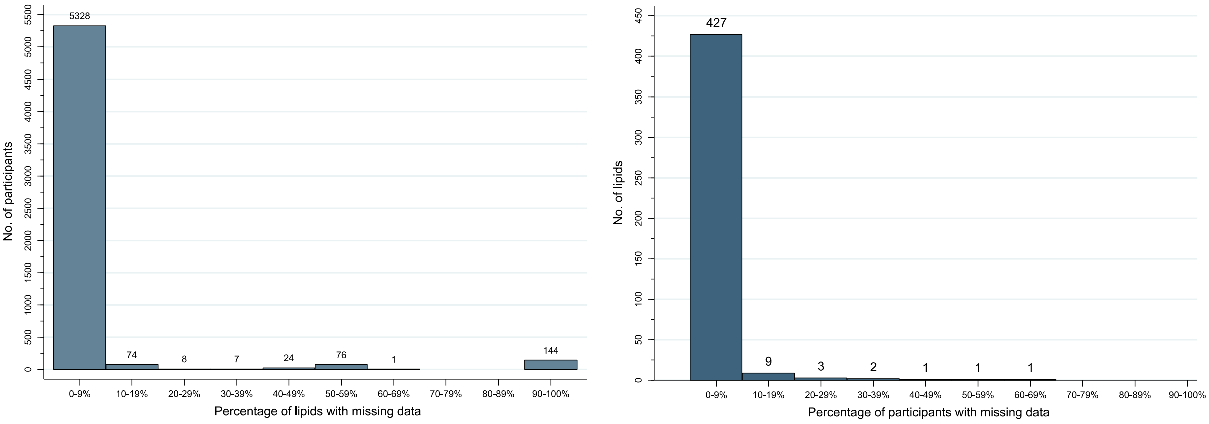
Fourth	rs1558861	chr11:116607437	T	C	1.0976	0.153	7.50E-13	*APOA5-APOC3*
Fourth	rs7483863	chr11:116652491	A	G	-1.082	0.152	9.10E-13	*APOA5-APOC3*
Fourth	rs71462009	chr11:116671824	T	C	-0.978	0.145	1.30E-11	*APOA5-APOC3*
Fourth	rs7350481	chr11:116586283	T	C	-0.95	0.141	1.60E-11	*APOA5-APOC3*
Fourth	rs180327	chr11:116623659	T	C	0.8018	0.121	3.50E-11	*APOA5-APOC3*
Fourth	rs11216140	chr11:116672013	T	C	-0.92	0.145	2.50E-10	*APOA5-APOC3*
Fourth	rs6589569	chr11:116671476	T	C	-0.914	0.145	2.70E-10	*APOA5-APOC3*
Fourth	rs9667814	chr11:116671823	C	G	-0.911	0.145	3.30E-10	*APOA5-APOC3*
Fourth	rs180326	chr11:116624703	T	G	0.7828	0.125	3.80E-10	*APOA5-APOC3*
Fourth	rs4938313	chr11:116671005	A	G	-0.894	0.144	5.00E-10	*APOA5-APOC3*
Fourth	rs6589567	chr11:116670676	A	C	-0.884	0.143	6.90E-10	*APOA5-APOC3*
Fourth	rs174568	chr11:61593816	T	C	-1.069	0.158	1.40E-11	*FADS1-2-3*
Fourth	rs174553	chr11:61575158	A	G	1.0427	0.156	2.60E-11	*FADS1-2-3*
Fourth	rs174545	chr11:61569306	C	G	1.0406	0.156	2.80E-11	*FADS1-2-3*
Fourth	rs174546	chr11:61569830	T	C	-1.04	0.156	2.80E-11	*FADS1-2-3*
Fourth	rs174547	chr11:61570783	T	C	1.0395	0.156	2.80E-11	*FADS1-2-3*
Fourth	rs174550	chr11:61571478	T	C	1.0395	0.156	2.80E-11	*FADS1-2-3*
Fourth	rs174554	chr11:61579463	A	G	1.0382	0.157	3.30E-11	*FADS1-2-3*
Fourth	rs174576	chr11:61603510	A	C	-1.025	0.155	3.50E-11	*FADS1-2-3*
Fourth	rs1535	chr11:61597972	A	G	1.0334	0.156	3.70E-11	*FADS1-2-3*
Fourth	rs35473591	chr11:61586328	CT	C	-1.03	0.157	5.10E-11	*FADS1-2-3*
Fourth	rs174562	chr11:61585144	A	G	1.0283	0.157	5.30E-11	*FADS1-2-3*
Fourth	chr11:61594920	chr11:61594920	CT	C	1.0243	0.157	6.40E-11	*FADS1-2-3*
Fourth	rs174583	chr11:61609750	T	C	-0.991	0.152	7.00E-11	*FADS1-2-3*
Fourth	chr11:61596322	chr11:61596322	CA	C	1.0201	0.157	7.50E-11	*FADS1-2-3*
Fourth	rs174580	chr11:61606642	A	G	1.01	0.155	7.90E-11	*FADS1-2-3*
Fourth	rs174581	chr11:61606683	A	G	-1.01	0.155	7.90E-11	*FADS1-2-3*
Fourth	rs174578	chr11:61605499	A	T	-1.005	0.155	9.60E-11	*FADS1-2-3*
Fourth	rs174577	chr11:61604814	A	C	-1.001	0.155	1.10E-10	*FADS1-2-3*
Fourth	rs174551	chr11:61573684	T	C	1.0622	0.165	1.10E-10	*FADS1-2-3*
Fourth	rs174564	chr11:61588305	A	G	1.0204	0.159	1.30E-10	*FADS1-2-3*
Fourth	rs174549	chr11:61571382	A	G	-1.038	0.164	2.40E-10	*FADS1-2-3*
Fourth	rs174567	chr11:61593005	A	G	0.9949	0.157	2.60E-10	*FADS1-2-3*
Fourth	rs174555	chr11:61579760	T	C	1.0359	0.164	2.80E-10	*FADS1-2-3*
Fourth	rs174536	chr11:61551927	A	C	0.9513	0.151	3.30E-10	*FADS1-2-3*
Fourth	chr11:61602460	chr11:61602460	CA	C	1.0333	0.165	3.50E-10	*FADS1-2-3*
Fourth	rs174566	chr11:61592362	A	G	0.981	0.156	3.50E-10	*FADS1-2-3*
Fourth	rs102274	chr11:61557826	T	C	0.9506	0.152	3.70E-10	*FADS1-2-3*
Fourth	rs174537	chr11:61552680	T	G	-0.943	0.151	4.10E-10	*FADS1-2-3*
Fourth	rs174561	chr11:61582708	T	C	1.021	0.164	5.20E-10	*FADS1-2-3*
Fourth	rs174544	chr11:61567753	A	C	-1.03	0.166	5.30E-10	*FADS1-2-3*
Fourth	rs99780	chr11:61596633	T	C	-0.972	0.157	5.30E-10	*FADS1-2-3*
Fourth	rs174556	chr11:61580635	T	C	-1.018	0.164	5.30E-10	*FADS1-2-3*
Fourth	rs174574	chr11:61600342	A	C	-0.965	0.156	5.40E-10	*FADS1-2-3*
Fourth	chr11:61602459	chr11:61602459	CCA	C	1.0396	0.168	5.60E-10	*FADS1-2-3*
Fourth	rs174533	chr11:61549025	A	G	-0.931	0.151	6.80E-10	*FADS1-2-3*

Fourth	rs174538	chr11:61560081	A	G	-0.993	0.162	7.60E-10	*FADS1-2-3*
Fourth	rs28456	chr11:61589481	A	G	1.0099	0.168	1.70E-09	*FADS1-2-3*
Fourth	rs174548	chr11:61571348	C	G	0.9823	0.163	1.70E-09	*FADS1-2-3*
Fourth	chr11:61591995	chr11:61591995	G	GAA	-0.99	0.165	1.80E-09	*FADS1-2-3*
Fourth	rs174535	chr11:61551356	T	C	0.8947	0.151	2.90E-09	*FADS1-2-3*
Fourth	rs174560	chr11:61581764	T	C	0.9703	0.164	3.00E-09	*FADS1-2-3*
Fourth	rs174534	chr11:61549458	A	G	0.9292	0.158	4.10E-09	*FADS1-2-3*
Fourth	rs102275	chr11:61557803	T	C	0.865	0.15	7.10E-09	*FADS1-2-3*
Fourth	rs174559	chr11:61581656	A	G	-0.989	0.171	7.70E-09	*FADS1-2-3*
Fourth	rs174530	chr11:61546592	A	G	0.8397	0.148	1.50E-08	*FADS1-2-3*

The association of the second, third, and fourth principal components of lipids with genetic variants is shown in order of *P*-value of significance within each principal component and genetic locus. The lead variant (SNP most strongly associated with each principal component) within each locus is highlighted. Abbreviations: EA = effect allele; GRCh37 = Genome Reference Consortium human genome build 37; NEA = non-effect allele; SE = standard error; SNP = single nucleotide polymorphism.

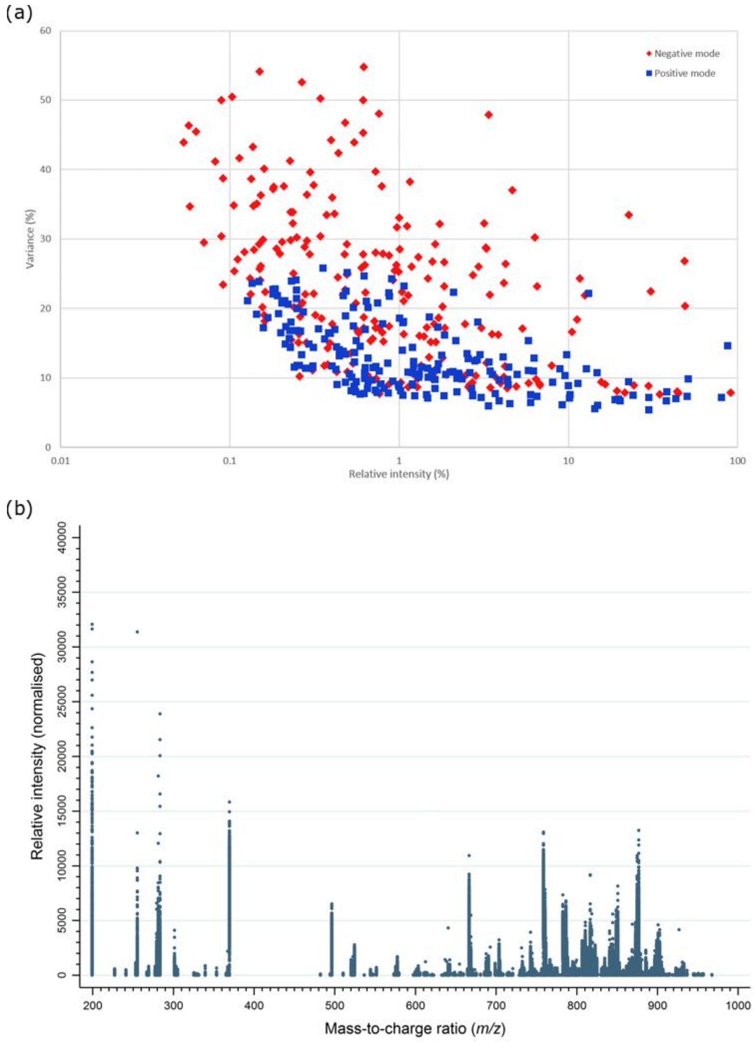
Supplementary Figure 1. Extent of missing data according to the number of lipids and participants

(a) (b)



(a) Percentage of lipids with missing data per participant. This subfigure shows that 5,328 out of 5,662 participants had less than 10% (0-9%) missing data, while 144 participants had 90-100% missing data. (b) Percentage of lipids with missing data per lipid. This subfigure shows that 427 out of 444 lipid peaks had less than 10% (0-9%) missing data, while there were 17 lipids that were missing in 10% or more of the lipids.

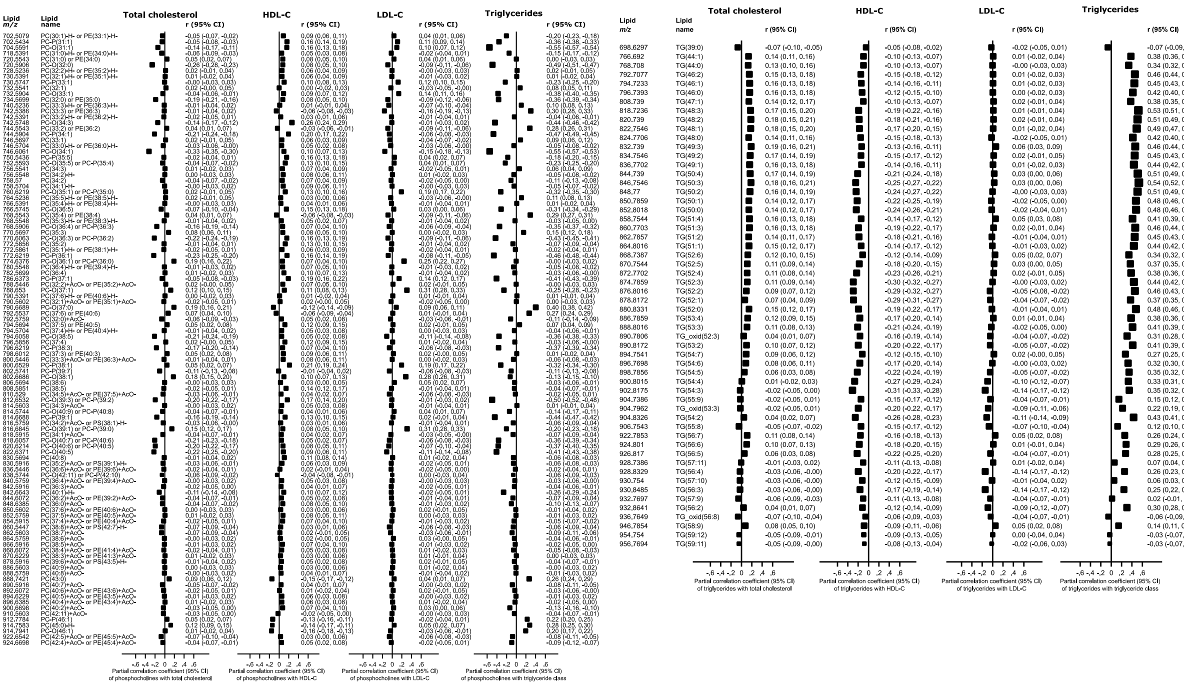
Supplementary Figure 2. Coefficients of variation and normalised relative intensities of lipids



(a) The coefficients of variation for each lipid expressed against the relative intensity for the quality control samples. (b) Normalised relative intensities of lipids.

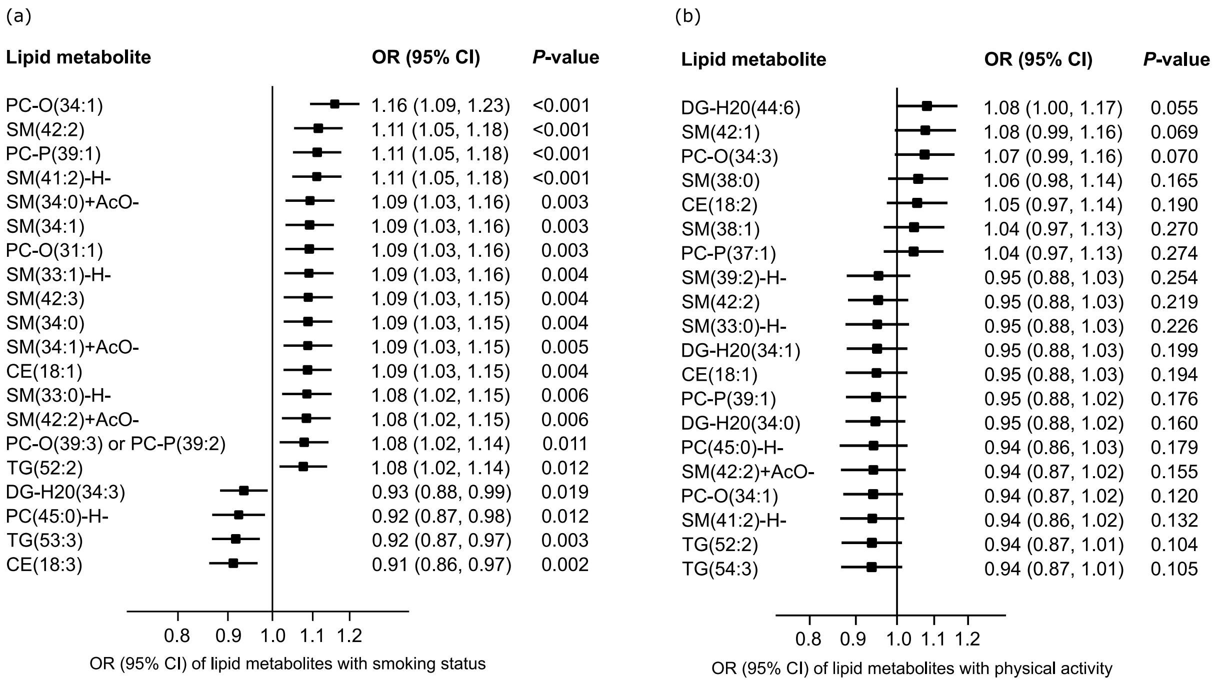
Supplementary Figure 3. Partial correlation coefficients of phosphatidylcholines and triacylglycerols with levels of major lipid markers

(a) **(b)**



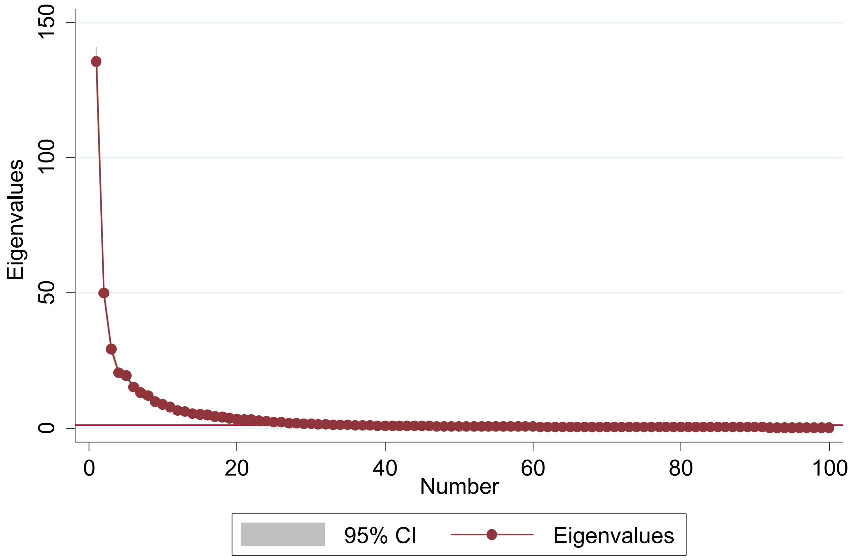
The partial correlations of (a) phosphatidylcholines and (b) triglycerides with total cholesterol, HDL-C, LDL-C, and total triacylglycerols are shown. All analyses were adjusted for age and sex.

Supplementary Figure 4. Association of lipids with smoking status and physical activity



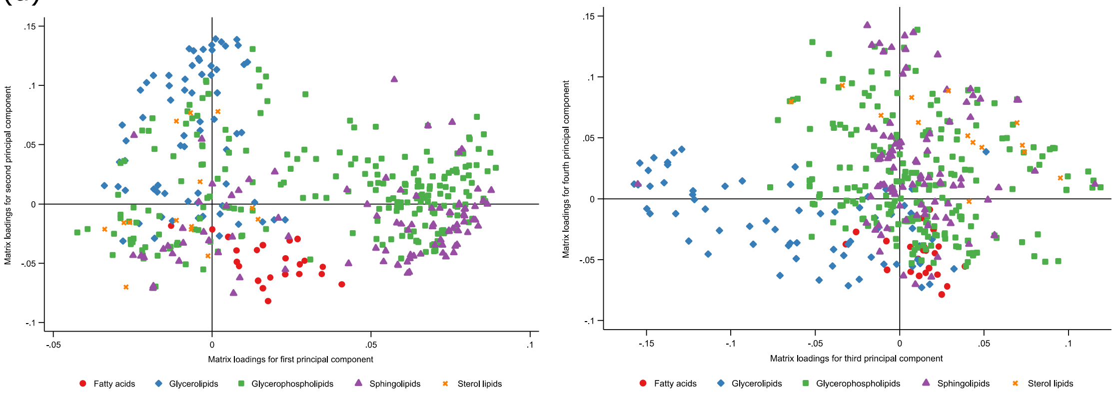
All analyses were adjusted for age and sex. Out of the lipids that were associated with rs662799 in the *APOA5-APOC3* locus, results are shown for (a) the top twenty lipids that were most significantly associated with smoking status and (b) the top twenty lipids that were most significantly associated with physical activity.

Supplementary Figure 5. Scree plot of eigenvalues from principal component analysis

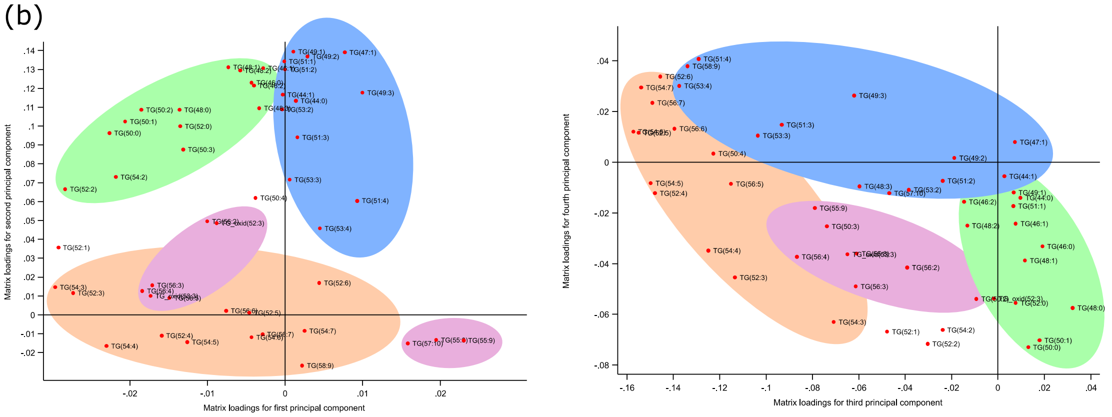


Supplementary Figure 6. Scatter plots showing matrix loadings of normalised relative intensities of lipids from principal component analysis

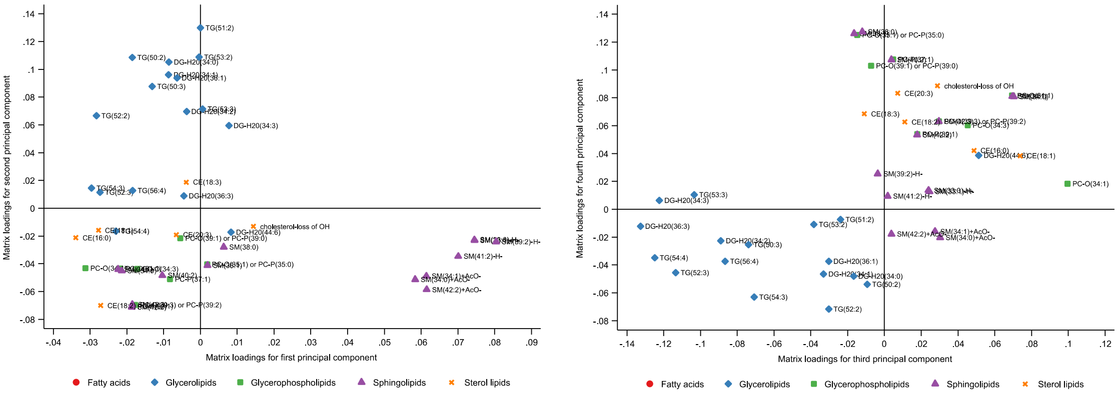
(a)



(b)

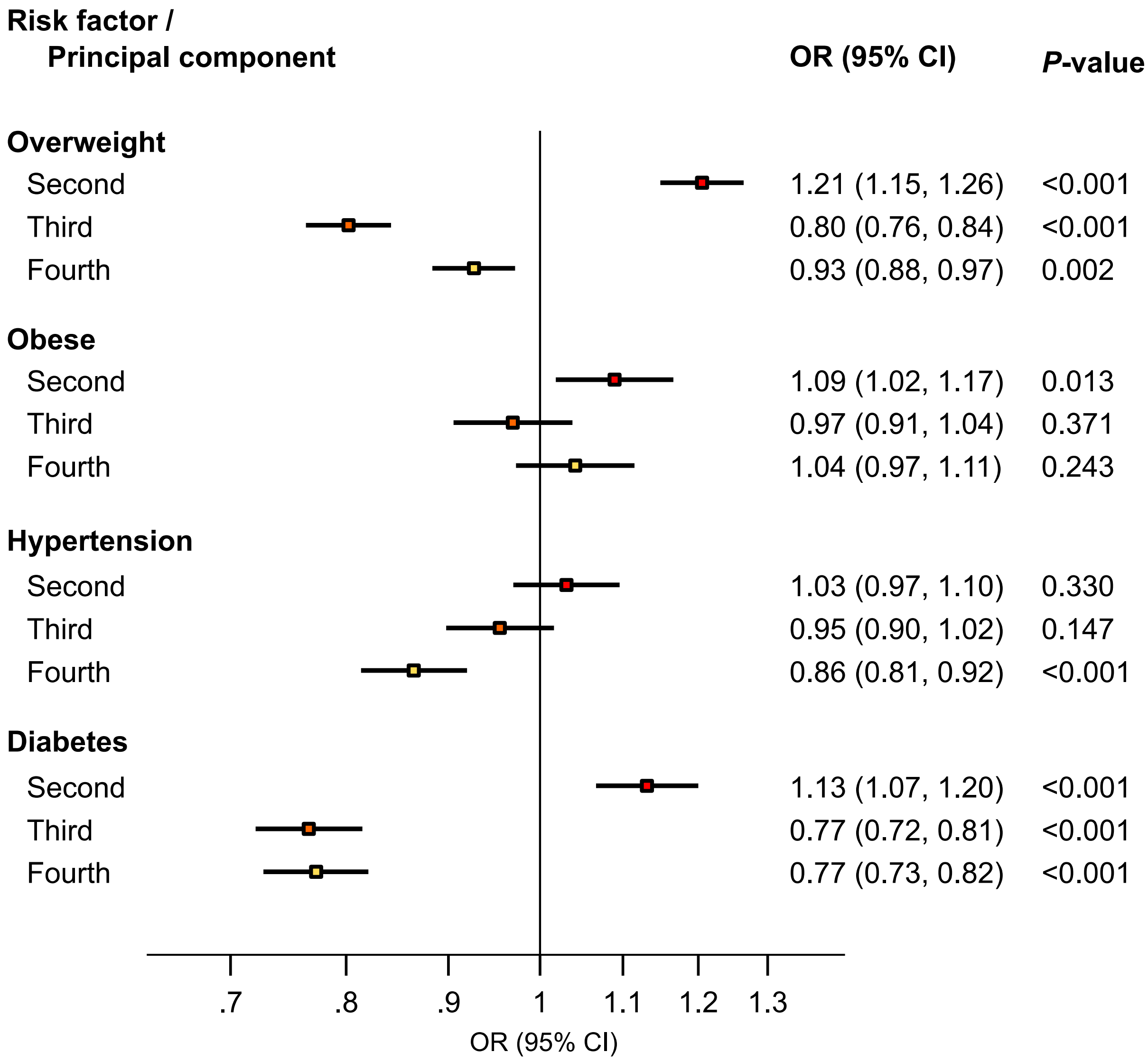


(c)



Lipids are coloured according to overall lipid category. (a) Overall lipids. (b) Subset of triacylglycerols. Blue oval indicates triacylglycerols containing odd-chain fatty acids increased by dairy consumption; orange oval indicates triacylglycerols containing ω-3 and ω-6 polyunsaturated fatty acids increased by fish consumption; green oval indicates triacylglycerols containing even-chain fatty acids formed in part through *de novo* lipogenesis. Pink ovals represent one or more distinct categories. (c) Subset of lipids associated with rs662799 in the *APOA5-APOC3* locus. Scatter plot is shown for the subset of lipids that are significantly associated with rs662799 (chr11:11663707) in the *APOA5-APOC3* locus at $P < 8.9 \times 10^{-10}$.

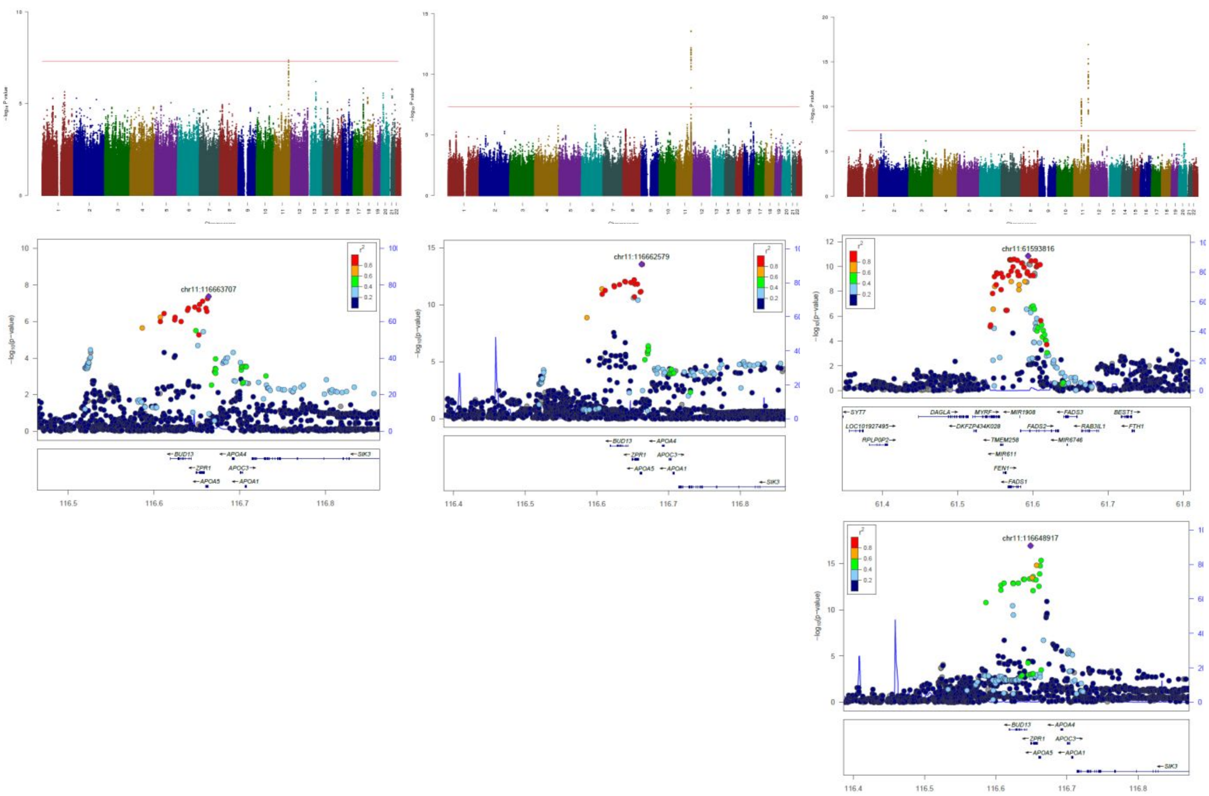
Supplementary Figure 7. Association of established coronary heart disease risk factors with principal components of lipid levels



All analyses were adjusted for age and sex. Odds ratios (OR) and 95% confidence intervals (CI) for each principal component are expressed per 1-SD increase in the loading scores of the lipids that make up that component. Definitions: Diabetes = $HbA_{1c} \geq 6.5\%$; Hypertension = SBP ≥ 140 mm Hg or DBP ≥ 90 mm Hg; Obese = BMI ≥ 30 kg/m^2; Overweight = BMI ≥ 25 kg/m^2. Abbreviations: BMI = body mass index; DBP = diastolic blood pressure; SBP = systolic blood pressure.

Supplementary Figure 8. Manhattan and regional association plots of principal components of lipid levels

(a) (b) (c)



Regional association plots were produced using LocusZoom. (a) Second principal component derived from lipid levels. (b) Third principal component derived from lipid levels. (c) Fourth principal component derived from lipid levels.