

1 **Appendix**

2

3 **Content**

4	Supplementary Note.....	2
5	References.....	6
6	Figure legends.....	7
7	Fig S1.....	10
8	Fig S2.....	11
9	Fig S3.....	12
10	Fig S4.....	13
11	Fig S5.....	14
12	Fig S6.....	15
13	Fig S7.....	16
14	Fig S8.....	17

Supplementary Note

Methods

Data sources

Genetic associations with the metabolites were sourced from Lotta *et al.* and the metabolites are grouped in the following classes: amino acids (AAs), biogenic amines (Bas), acylcarnitines (ACs), phosphatidylcholines (PCs), lysophosphatidylcholines (LPCs), sphingomyelins (SMs), and hexoses.

Genetic associations (i.e., point estimates and standard errors) with clinical outcomes were obtained from genome-wide association studies on atrial fibrillation (AF), heart failure (HF), dilated cardiomyopathy (DCM), and non-ischemic cardiomyopathy (NICM). Below, we summarize the study-specific characteristics.

- AF from Nielsen *et al.*(1), enrolling 1,030,836 European participants, including 60,620 cases. This meta-analysis included five prospective studies with study-specific definitions of AF:
 - ICD-10/9 codes I48 and 427.3 for the HUNT study.
 - ICD-9 code 427.31 for the Michigan Genomics Initiative.
 - ICD-10 code I48 for the DiscovEHR Collaboration Cohort.
 - ICD-10/9 codes I48 and 427.3 for the UK biobank.
 - A combination of EHR diagnosis (e.g., using ICD codes) and ECG/Holter measurements for the Christophersen *et al.* 2017 GWAS.
- HF from Shah *et al.*(2), enrolling 977,323 participants, including 47,309 cases. This meta-analysis of 26 studies defined HF a clinical diagnosis of HF of any aetiology with no inclusion criteria based on LV ejection fraction. The studies involved both case-control studies as well as population-based studies.

- DCM from Garnier *et al.*(3), enrolling 7,159 participants, including 2,719 cases. This meta-analysis of nine studies excluded the UK biobank and defined DCM based on the presence of reduced ejection fraction and an enlarged left ventricular end-diastolic volume/diameter in the absence of significant coronary artery disease, intrinsic valvular disease, documented myocarditis, systemic disease (such as sarcoidosis), sustained arterial hypertension, or congenital malformation.
- NICM from Aragam *et al.*(4), enrolling 395,972 participants, of whom 1,816 cases. This GWAS was exclusively conducted in the UK biobank, with NICM cases defined as an HF diagnosis with LV dysfunction in the absence of CAD. LV dysfunction was defined as a diagnosis of DCM or LV failure. CAD was defined as a diagnosis of myocardial infarction or having received a coronary revascularization.

Genetic associations (i.e., point estimates and standard errors) with protein levels were obtained from eight genome-wide association studies, which are summarized here:

- deCODE(5) measured plasma protein levels in 35,559 participants using a SomaLogic assay of 4,907 aptamers. The participants were mainly obtained by the Icelandic Cancer Project and the population-based deCODE Health study.
- The meta-analysis of SCALLOP(6) in 21,758 participants from 13 cohorts of European ancestry and 9,173 participants from two replication cohorts comprised the Olink CVD-I proteins.
- Ahola-Olli *et al.*(7) included 8,293 participants from three independent population-based Finnish studies. *et al.* and measured cytokine levels using BioRad assays.
- Framingham(8) measured 85 plasma protein levels using a Luminex assay in 6,861 participants of the prospective community-based FHS Offspring and Third Generation cohorts.
- The AGES-Reykjavik(9) study consisted of a random population of 5,368 participants living in Reykjavik in 1967 and their plasma protein levels were measured using 4,782 aptamers of the SomaLogic assay.

- The INTERVAL(10) substudy consisted of 3,301 blood donors from England and measured plasma protein levels using 4,034 aptamers of the SomaLogic assay.
- Gilly *et al.*(11) measured plasma protein levels of the CVDII, CVDIII and Metabolism Olink assays in 1,328 participants.
- Yang *et al.*(12) included 636 participants and measured plasma protein levels using 1,305 aptamers of the SomaLogic assay.

mRNA expression and enrichment

Cardiac mRNA expression was obtained from the human protein atlas (HPA), sourcing the consensus expression obtained by normalised transcripts per million (nTPM) values from three independent transcriptomics datasets: GTEx, Fantom5, and HPA's own data. Overexpressed genes were identified by comparing cardiac expression with average expression in other tissues, testing against a standard normal quantile of 1.96. Enrichment analysis, evaluating the frequency of association across metabolite classes was conducted using Fisher's exact.

Networks

We identified metabolites clusters based on a shared protein profile (defined as 20% or more protein in common) and visualised the triangulated associations using annotated networks. Reactome pathway enrichment of the proteins belonging to the metabolite networks compared to the remaining set of proteins was tested using a Wald test.

Results

Replicating protein associations with metabolites

Out of the 82 prioritised proteins, 49 were available in more than one GWAS, allowing for replication of the associations with plasma metabolites. Applying a nominal replication p-value of 0.05 we were able to replicate the metabolite association of 45 proteins (91.8%; **Table S13**). Applying a more stringent p-value cut-off resulted in 38 replicated proteins.

1

2 *Networks*

3 Cluster analysis was performed to identify plasma metabolites sharing common protein
4 effects, identifying four groups (**Figure 1C, Fig S4-7**). This highlighted the following
5 biological pathways: “TP53 regulates transcription of cell death genes” and “metabolism of
6 RNA” (**Fig S8**).

7

References

1. Nielsen JB, Thorolfsson RB, Fritsche LG, Zhou W, Skov MW, Graham SE, et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet.* 2018;50(9):1234–9.
2. Shah S, Henry A, Roselli C, Lin H, Sveinbjörnsson G, Fatemifar G, et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun.* 2020;11(1):1–12.
3. Garnier S, Harakalova M, Weiss S, Mokry M, Regitz-Zagrosek V, Hengstenberg C, et al. Genome-wide association analysis in dilated cardiomyopathy reveals two new players in systolic heart failure on chromosomes 3p25. 1 and 22q11. 23. *Eur Heart J.* 2021;42(20):2000–11.
4. Aragam KG, Chaffin M, Levinson RT, McDermott G, Choi SH, Shoemaker MB, et al. Phenotypic refinement of heart failure in a national biobank facilitates genetic discovery. *Circulation.* 2019;139(4):489–501.
5. Ferkingstad E, Sulem P, Atlason BA, Sveinbjörnsson G, Magnusson MI, Styrismisdóttir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet.* 2021;53(12):1712–21.
6. Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman ÅK, Schork A, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab.* 2020;2(10):1135–48.
7. Ahola-Olli AV, Würtz P, Havulinna AS, Aalto K, Pitkänen N, Lehtimäki T, et al. Genome-wide association study identifies 27 loci influencing concentrations of circulating cytokines and growth factors. *Am J Hum Genet.* 2017;100(1):40–50.
8. Yao C, Chen G, Song C, Keefe J, Mendelson M, Huan T, et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat Commun.* 2018;9(1):3268.
9. Gudjonsson A, Gudmundsdóttir V, Axelsson GT, Gudmundsson EF, Jonsson BG, Launer LJ, et al. A genome-wide association study of serum proteins reveals shared loci with common diseases. *Nat Commun.* 2022;13(1):480.
10. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. *Nature.* 2018;558(7708):73–9.
11. Gilly A, Park YC, Png G, Barysenka A, Fischer I, Bjørnland T, et al. Whole-genome sequencing analysis of the cardiometabolic proteome. *Nat Commun.* 2020;11(1):6336.
12. Yang C, Farias FH, Ibanez L, Suhy A, Sadler B, Fernandez MV, et al. Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. *Nat Neurosci.* 2021;24(9):1302–12.

Figure legends

Fig S1. Volcano plots displaying proteins associated with the metabolites.

NB. Labelled proteins are drugged, which is defined as proteins targeted by a compound (see Methods); the p-value was truncated to a -log₁₀ of 16 for visualisation purposes only.

Abbreviations: a = acyl residue, aa = diacyl residue, ae = acyl-alkyl residue, LPC = lysophosphatidylcholine, MD = mean difference, PC = phosphatidylcholine, SM = sphingomyelin.

Fig S2. Volcano plots displaying proteins associated with the cardiac outcomes.

NB. Labelled proteins are drugged, which is defined as proteins targeted by a compound (see Methods); the p-value was truncated to a -log₁₀ of 16 for visualisation purposes only.

Abbreviations: AF = atrial fibrillation, DCM = dilated cardiomyopathy, HF = heart failure, NICM = non-ischemic cardiomyopathy, OR = odds ratio.

Fig S3. Enriched Reactome pathways and their corresponding -log₁₀(p-value) per cardiac outcome.

Figure S4. Heatmap displaying the percentage of overlapping associated proteins between metabolite pairs.

NB. Orange boxes represent clusters further studied in detail.

Abbreviations: a = acyl residue, aa = diacyl residue, ae = acyl-alkyl residue, LPC = lysophosphatidylcholine, PC = phosphatidylcholine, SM = sphingomyelin.

Fig S5. Annotated network of prioritised metabolites, proteins, and outcomes for which the metabolites have at least 20% associated common proteins belonging to the metabolite classes LPC and PC.

NB. Prioritised proteins are represented by circles, metabolites by diamonds, outcomes by triangles. Circle colours represent protein druggability, where drugged is defined as targeted by an approved compound and druggable as targeted by a compound, see Methods. Increasing effect is displayed by a red arrow, decreasing effect by a blue arrow. Abbreviations: a = acyl residue, aa = diacyl residue, AF = atrial fibrillation, HF = heart failure, LPC = lysophosphatidylcholine, PC = phosphatidylcholine.

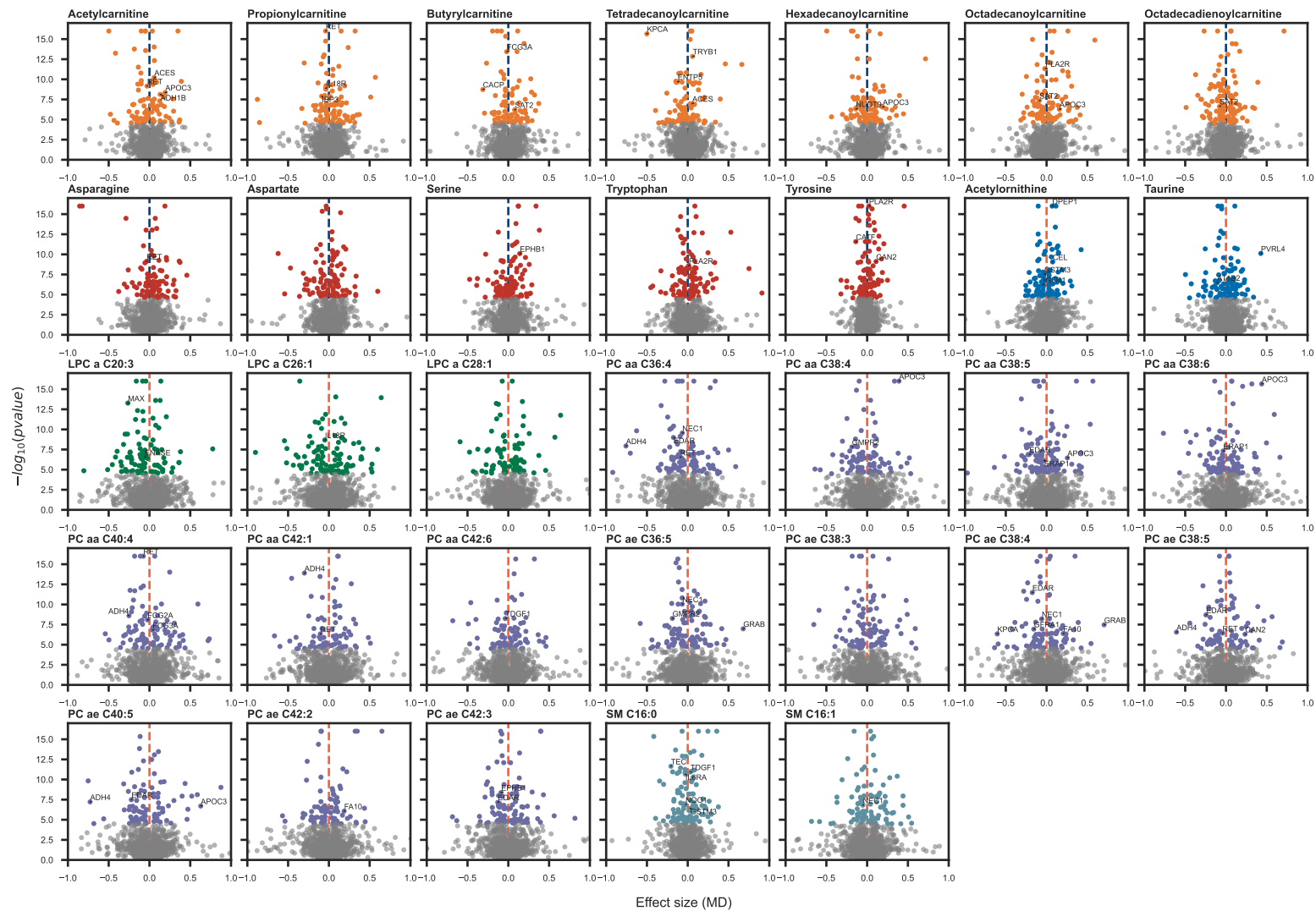
Fig S6. Annotated network of prioritised metabolites, proteins, and outcomes for which the metabolites have at least 20% associated common proteins belonging to the metabolite class PC.

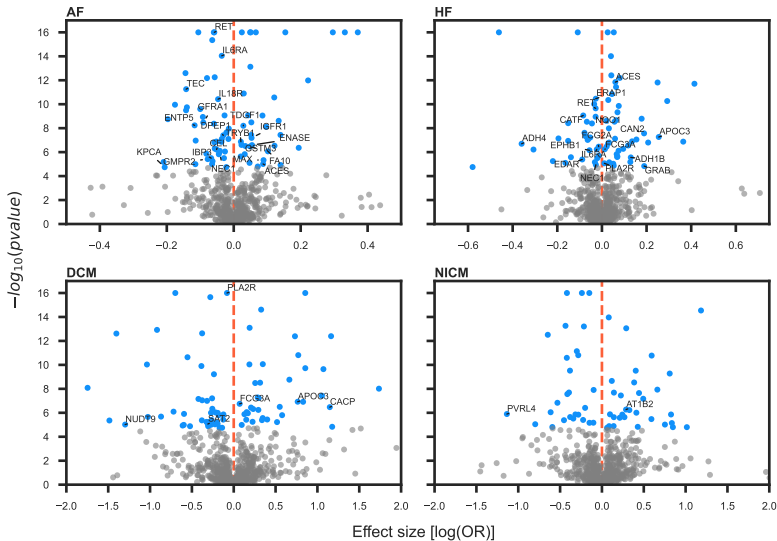
NB. Prioritised proteins are represented by circles, metabolites by diamonds, outcomes by triangles. Circle colours represent protein druggability, where drugged is defined as targeted by an approved compound and druggable as targeted by a compound, see Methods. Increasing effect is displayed by a red arrow, decreasing effect by a blue arrow. Abbreviations: ae = acyl-alkyl residue, AF = atrial fibrillation, HF = heart failure, PC = phosphatidylcholine.

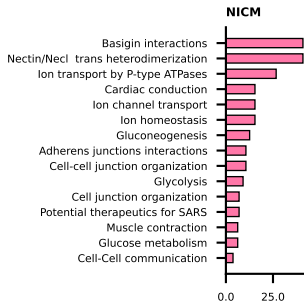
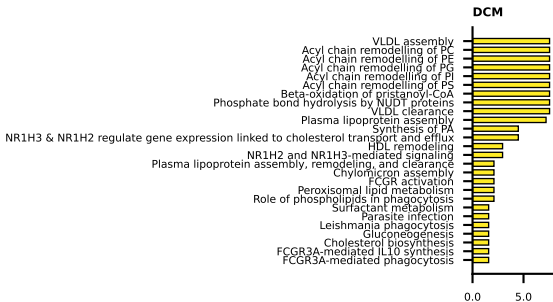
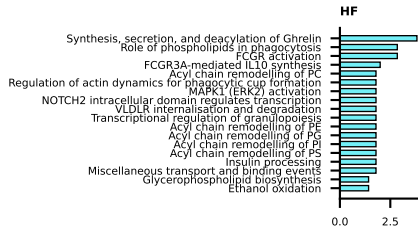
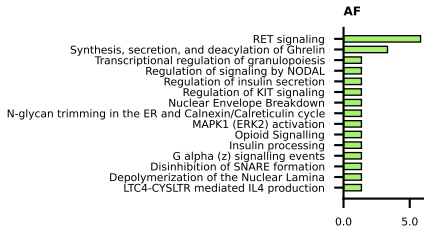
Fig S7. Annotated network of prioritised metabolites, proteins, and outcomes for which the metabolites have at least 20% associated common proteins belonging to the metabolite classes LPC and PC.

NB. Prioritised proteins are represented by circles, metabolites by diamonds, outcomes by triangles. Circle colours represent protein druggability, where drugged is defined as targeted by an approved compound and druggable as targeted by a compound, see Methods. Increasing effect is displayed by a red arrow, decreasing effect by a blue arrow. Abbreviations: a = acyl residue, aa = diacyl residue, ae = acyl-alkyl residue, AF = atrial fibrillation, HF = heart failure, LPC = lysophosphatidylcholine, PC = phosphatidylcholine.

- 1 **Fig S8. Enriched Reactome pathways and their corresponding $-\log_{10}(\text{p-value})$ per**
- 2 **metabolite cluster.**
- 3 CAS = contact activation system (CAS); KKS = kallikrein/kinin system
- 4 Abbreviations: LPC = lysophosphatidylcholine, PC = phosphatidylcholine.







$-\log_{10}(p)$

