Association analyses based on false discovery rate implicate many new loci for coronary artery disease

Christopher P. Nelson[1,2,61], Anuj Goel[3,4,61], Adam S Butterworth[5,6,61], Stavroula Kanoni[7,8,61], Tom R. Webb[1,2], Eirini Marouli[7,8], Lingyao Zeng[9], Ioanna Ntalla[7,8], Florence Y. Lai[1,2], Jemma C. Hopewell[10], Olga Giannakopoulou[7,8], Tao Jiang[5], Stephen E. Hamby[1,2], Emanuele Di Angelantonio[5,6], Themistocles L. Assimes[11], Erwin P. Bottinger[12], John C. Chambers[13,14,15], Robert Clarke[10], Colin NA Palmer[16,17], Richard M. Cubbon[18], Patrick Ellinor[19], Raili Ermel[20], Evangelos Evangelou[21,22], Paul W. Franks[23,24,25], Christopher Grace[3,4], Dongfeng Gu[26], Aroon D. Hingorani[27], Joanna M. M. Howson[5], Erik Ingelsson[28], Adnan Kastrati[9], Thorsten Kessler[9], Theodosios Kyriakou[3,4], Terho Lehtimäki[29], Xiangfeng Lu[26], Yingchang Lu[12,30], Winfried März[31,32,33], Ruth McPherson[34], Andres Metspalu[35], Mar Pujades-Rodriguez[36], Arno Ruusalepp[20,37], Eric E. Schadt[38], Amand F. Schmidt[39], Michael J. Sweeting[5], Pierre A. Zalloua[40,41], Kamal AlGhalayini[42], Bernard D. Keavney[43,44], Jaspal S. Kooner[14,15,45], Ruth J. F. Loos[12,46], Riyaz S. Patel[47,48], Martin K. Rutter[49,50], Maciej Tomaszewski[51,52], Ioanna Tzoulaki[21,22], Eleftheria Zeggini[53], Jeanette Erdmann[54,55,56], George Dedoussis[57], Johan L. M. Björkegren[37,38,58], EPIC-CVD Consortium, CARDIoGRAMplusC4D, The UK Biobank CardioMetabolic Consortium CHD working group, Heribert Schunkert[9,61], Martin Farrall[3,4,61], John Danesh[5,6,59,61], Nilesh J. Samani[1,2,61], Hugh Watkins[3,4,61], Panos Deloukas[7,8,60,61]

1. Department of Cardiovascular Sciences, University of Leicester, Leicester LE3 9QP, UK
2. National Institute for Health Research Leicester Biomedical Research Centre, Leicester LE3 9QP, UK
3. Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DU, UK
4. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK
5. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
6. NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
7. William Harvey Research Institute, Barts & the London Medical School, Queen Mary University of London, London EC1M 6BQ, UK
8. Centre for Genomic Health, Queen Mary University of London, London EC1M 6BQ, UK
9. German Heart Center Munich, Clinic at Technische Universität München and Deutsches Zentrum für Herz- und Kreislauferkrankungen (DZHK), partner site Munich Heart Alliance, Munich 80636, Germany
10. CTSU, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK
11. Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA
12. Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, NewYork City,NY 10029, USA
13. Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK
14. Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, Middlesex UB1 3HW, UK
15. Imperial College Healthcare NHS Trust, London W12 0HS, UK
16. Molecular and Clinical Medicine, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee, UK
17. Pharmacogenomics Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee
18. Leeds Institute of Cardiovascular and Metabolic Medicine, The University of Leeds, Leeds LS2 9JT, UK
19. Cardiac Arrhythmia Service and Cardiovascular Research Center, The Broad Institute of Harvard and MIT, Boston, USA
20. Department of Cardiac Surgery, Tartu University Hospital, Tartu 50406, Estonia
21. Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK
22. Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina 45110, Greece
23. Department of Clinical Sciences, Genetic & Molecular Epidemiology Unit, Lund University Diabetes Center, Skåne University Hospital, Lund University, Malmö/Skåne SE-205 02, Sweden
24. Department of Nutrition, Harvard T. H. Chan School of Public Health, Harvard University, Boston/Massachusetts MA 02115, USA

25. Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, Umeå/Västerbotten SE-901 85, Sweden
26. State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100037, China
27. University College London, London, UK
28. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, Stanford, CA 94305, US
29. Department of Clinical Chemistry, Fimlab Laboratories and Faculty of Medicine and Life Sciences, University of Tampere, Tampere 33520, Finland
30. Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN 37203, USA
31. Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz 8036, Austria
32. Medical Clinic V (Nephrology, Rheumatology, Hypertensiology, Endocrinology, Diabetology), Medical Faculty Mannheim, University of Heidelberg, Mannheim 68167, Germany
33. Academy, Synlab Holding Deutschland GmbH, Mannheim 68161, Germany
34. Ruddy Canadian Cardiovascular Genetics Centre University of Ottawa Heart Institute, Ottawa K1Y 4W7, Canada
35. Estonian Genome Center, University of Tartu, Tartu 51010, Estonia
36. Leeds Institute of Biomedical and Clinical Sciences, The University of Leeds, Leeds LS2 9JT, UK
37. Clinical Gene Networks AB, Stockholm 114 44, Sweden
38. Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, NYC, NY 10029-6574, USA
39. Institute of cardiovascular science, UCL, London, UK
40. Lebanese American University, School of Medicine, Beirut 13-5053, Lebanon
41. Harvard T.H. Chan School of Public Health, Boston Zip 02115, USA
42. Department of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia
43. Division of Cardiovascular Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester M13 9PT, UK
44. Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 9WL, UK
45. Cardiovascular Science, National Heart and Lung Institute, Imperial College London, London W12 0NN, UK
46. Mindich Child Health Development Institute, Icahn Shool of Medicine at Mount Sinai, New York, NY 10069, USA, NewYork City,NY 10029, USA
47. Farr Institute of Health Informatics, UCL, London NW1 2DA, UK
48. Bart's Heart Centre, St Bartholomew's Hospital, London Ec1A 7BA, UK
49. Division of Diabetes, Endocrinology and Gastroenterology, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester M13 9PT, UK
50. Manchester Diabetes Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 0JE, UK
51. Division of Cardiovascular Sciences Faculty of Biology, Medicine and Health The University of Manchester, Manchester M13 9PT, UK
52. Division of Medicine, Central Manchester NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 9PT, UK
53. Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK
54. Institute for Cardiogenetics, University of Lübeck, Lübeck 23562, Germany
55. DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, Lübeck 23562, Germany
56. University Heart Center Luebeck, Lübeck 23562, Germany
57. Department of Nutrition-Dietetics/Harokopio University, Athens, Greece
58. Integrated Cardio Metabolic Centre, Department of Medicine, Karolinska Institutet, Karolinska Universitetssjukhuset, Huddinge 141 57, Sweden
59. Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1RQ, UK
60. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia
61. These authors contributed equally to this work

Correspondence should be addressed to: Hugh Watkins (hugh.watkins@rdm.ox.ac.uk**)**

Genome-wide association studies (GWAS) in coronary artery disease (CAD) have identified 66 loci at 'genome-wide significance' ($p < 5 \times 10^{-8}$) but a much larger number of putative loci at a false discovery rate (FDR) of 5%[1-4]. Here, we leverage an interim release of UK Biobank (UKBB) data to evaluate the validity of the FDR approach. We tested a CAD phenotype inclusive of angina (SOFT; N_{cases}=10,801) as well as a stricter definition without it (HARD; N_{cases}=6,482) and selected the former for conducting a meta-analysis with the two most recent CAD GWASs[2-3]. This approach identified 13 new loci at genome-wide significance, 12 of which were in our previous 5% FDR list[2], and provided strong support that the remaining FDR loci represent genuine signals. The set of 304 independent variants at 5% FDR in this study explain 21.2% of CAD heritability and identified 243 loci that implicate pathways in blood vessel morphogenesis as well as lipid metabolism, nitric oxide signaling and inflammation.

Previous GWAS studies of CAD risk[1-4] have interrogated a large number of cases and controls but remain less well-powered than GWAS of quantitative traits[5]. UKBB was established to improve understanding of the causes of common diseases including CAD, a leading health problem around the world[6]. In addition to self-reported disease outcomes and extensive health and life-style questionnaire data, the 502,713 participants are being tracked through their NHS records and national registries (including cause of death and Hospital Episode Statistics). In July 2015, UKBB released genotypes imputed to the 1000 Genomes panel for 152,249 participants profiled with a SNP array harboring 820,967 variants comprising common variants optimized for imputation, validated rare coding variants and sets of phenotype-associated variants or their proxies (e.g. GWAS catalogue).

We set up The UKBiobank-CardioMetabolic-Consortium CHD working group to assess the use of self-reported and hospital record data on CAD in UKBB and define the relevant case and control subgroups to undertake genetic analyses of CAD risk.

The July 2015 release of UKBB comprises 10,801 genotyped individuals with an inclusive CAD phenotype ('SOFT') that incorporates self-reported angina or other evidence of chronic coronary heart disease, of which 6,482 have a more stringently defined CAD phenotype ('HARD') of myocardial infarction and/or revascularisation (**Fig. 1a**). After QC we analysed the SOFT and HARD cases separately against 137,914 controls for 9,149,595 variants present either in the CARDIoGRAMplusC4D 1000-Genomes GWAS[2] or the MIGen/CARDIoGRAM Exome-chip study[3-4]. The SOFT definition was selected for the primary analysis based on power calculations (**Supplementary Table 1**). We found 4 (SOFT and HARD), 1 (SOFT only) and 2 (HARD only) variants reaching genome-wide significance, all located in known CAD loci (**Supplementary Figure 1**).

We then meta-analysed the UKBB data for each CAD definition with each of the two published data sets (**Supplementary Figure 2**) using an inverse-variance weighted fixed-effect (IVW-FE) model and double genomic control correction (**Online Methods**). For both the SOFT and HARD definitions, we validated all 66 known CAD loci (72 independent variants with $p < 1.2 \times 10^{-3}$ IVW-FE) with 43 and 37 respectively reaching genome-wide significance in this study (**Supplementary Table 2**). Outside the known CAD loci (1 Mb window centred on the published lead SNP) we found 9 new signals (in both SOFT and HARD) reaching genome-wide significance (**Table 1** and **Fig. 2**). The anticipated increase in power with the SOFT definition (**Supplementary Table 1**) was attenuated by an inflation of the lambda statistic (**Supplementary Table 3**), potentially due to a combination of larger sample size (i.e. polygenicity) and a less homogeneous phenotype in the SOFT definition. Overall, there was

strong concordance between corresponding signals for SOFT and HARD (**Fig. 1b, Supplementary Table 4**); subsequent analyses were undertaken using the SOFT meta-analysis results.

To look for additional signals beyond the 9 that reached genome-wide significance (**Fig. 2**) we performed an FDR analysis and selected 23 suggestive signals at 1% FDR ($p < 1.55x10^{-6}$ IVW-FE; **Supplementary Table 4**) outside known CAD loci which we validated in an independent sample of up to 4,412 cases and 3,910 controls from the German MI-Family-Studies V and VI and a Greek case-control study (**Supplementary Table 5**). In total, we identified 13 new genome-wide significant CAD loci in the combined discovery and replication sample (**Table 1, Supplementary Table 6**).

In our recent large-scale GWAS[2], we reported 162, mainly common, variants at an FDR discovery cutoff of 5% showing conditional independent associations with the P_{joint} test in GCTA[7]. Twelve of the 13 new sentinel SNPs were present or had a proxy ($r^2>0.8$) among these 162 variants[2]. **Fig. 3** shows a strong linear relationship between association signals for these 162 variants in the earlier[2] and current analysis, with overall greater significance levels in the current meta-analysis. As expected, we observed an excess of small p-values for this set of variants in the UK Biobank alone (**Supplementary Figure 3a**). Monte Carlo simulations show that the expected number of replicated variants in the UK Biobank data is 56 (95% CI 42 – 69) (**Supplementary Figure 3b**) and we found 58 variants after allowing for multiple testing (q-values < 0.05). This further confirms the validity of extended lists of associated variants based on FDR criteria. We therefore defined a new FDR list of association signals by performing an approximate joint association analysis with the GCTA software[7] as described elsewhere[2] using the 11,427 SNPs with 5%FDR. We identified 304 independent variants at $P_{joint} < 10^{-4}$, clustering in 243 putative CAD loci (**Supplementary Table 7**). The new 5%FDR set overlaps by 122 SNPs

with the old set (75.3%; including proxies at an $r^2 > 0.8$). We then assessed heritability using the independent set of 304 SNPs and obtained a heritability estimate of 21.2%. The contribution to this heritability estimate of the 13 new loci (**Table 1**) was 1.03% whereas the new and known genome-wide significant CAD loci together explained 8.53% of CAD heritability. To further assess the validity and utility of the 5%FDR set, we tested the ability to predict CAD using genetic risk scores (GRS) based on either the 5%FDR SNPs (GRS1) or only CAD variants reaching genome-wide significance (GRS2; **Online Methods**) in an independent sample, EPIC-CVD[8], comprising 7910 CHD cases and 12958 controls. In a model with age and sex, GRS1 increased the C-index by 0.25% compared to GRS2 (**Supplementary Table 8**). GRS1 improved the point estimates of the HR compared to GRS2 mainly in the second (from 0.9116 to 0.8314) and fourth quintile (from 1.0437 to 1.176), **Supplementary Figure 4**.

We then explored the biology of the 13 new genome-wide significant CAD risk loci; **Supplementary Figure 5** shows regional association plots**. Supplementary Figure 6** provides *in silico* functional annotation (**Online Methods**) for each lead variant and its proxies (1000 Genomes). We found compelling evidence to implicate candidate genes *ITGB5*, *TGB1*, *PDE5A*, *ARHGEF26*, *FN1*, *CDH13*, and *HNF1* (detailed in **Supplementary Note)**. The risk allele of rs150512726 (proxy for rs142695226; **Table 1**), causes a 3 amino acid deletion within the cytoplasmic tail of integrin subunit beta 5 (ITGB5), part of a heterodimer which regulates the activation of latent TGFB1 (Transforming growth factor beta 1)[9-10]. The intronic variant (rs8108632; **Table 1**) we identified in *TGFB1*, further implicates the TGFB1 pathway in CAD risk. TGFB1 is known to have important roles in endothelium and vascular smooth muscle[11] but has not been widely studied in atherosclerosis, though a recent study implicates TGF-β signalling downstream of CDKN2B in the *CDKN2BAS* cardiovascular risk locus[12]. eQTL analyses suggested candidate CAD risk genes (*TDRKH*, *FN1*, *ARHGEF26*, *PDE5A*, *ARNTL*, and *CDH13*) in

six new loci (**Supplementary Table 9**). For example, the lead variant rs7678555 (**Table 1**) was found to be a strong eQTL ($p=8.1x10^{-13}$ linear regression model) for *PDE5A* only in aorta from CAD patients (STARNET[13]; **Supplementary Table 9**) although its regulatory potential was modest using functional prediction tools (**Online methods**). *PDE5A* encodes a cGMP-specific phosphodiesterase which is important for smooth muscle relaxation in the cardiovascular system where it regulates nitric-oxide-generated cGMP[14]. Furthermore, mining eQTL data in tissues from CAD patients (STARNET) showed several other instances of eSNPs (*TDRKH*, FN1, *CDH13*; **Supplementary Table 9**) having no effect in tissues from non-CAD patients (GTEx[15]). One caveat is that sample size differs between STARNET and GTex for certain tissues. Nonetheless, our observation highlights the need to expand efforts to map regulatory elements in disease tissues.

Other candidate genes fit with emerging data on atherosclerosis mechanisms. For example, a knockout mouse for *ARHGEF26* on a hyperlipidemic background resulted in reduced atherosclerosis and plaques with reduced macrophage content[16]. Similarly, *FN1* expression is increased in plaques and mouse models have demonstrated a causal role for fibronectin-1 in the development and progression of atherosclerosis[17-18]. Finally, we undertook a phenome scan to assess pleiotropy (**Supplementary Table 10**). Several of the new lead SNPs (or a proxy) had robust associations ($p < 5x10^{-8}$ meta-analysis) with traditional CAD risk factors such as LDL-cholesterol (*HNF1A* and *FN1*), blood pressure (*PRDM8/FGF5*) and BMI (*SNRPD2*).

We next evaluated the broader functional relationships among genes associated with variants (N=11,427) at 5%FDR. The 5%FDR set was annotated for eQTLs which, when present, were mainly found in atherosclerotic aortic wall (25%) or internal mammary artery (22%) of CAD

patients (STARNET[13]; **Supplementary Table 9**). In GTEx[15], eQTLs were mainly found in subcutaneous fat (**Supplementary Table 9; Supplementary Figure 7**).

Prior pathway analyses of GWAS CAD loci have highlighted genes involved in lipid metabolism, cellular movement, and processes such as tissue morphology and immune cell trafficking[1]. Analysis of 357 genes, selected as either eQTLs and/or the nearest gene to a 5%FDR independent variant in this study (N=304), with the Ingenuity Knowledge base confirmed the above findings[1] highlighting cardiovascular system development and function ($p = 1.31x10^{-16}$ right-tailed Fisher Exact Test (rtFET)), organismal development ($p = 1.31 \times 10^{-16}$ rtFET) and survival ($p = 1.52x10^{-16}$ rtFET) as the most significant processes. In addition to canonical pathways related to lipid metabolism, extracellular matrix, inflammation and nitric oxide production, the 357 gene set showed enrichment for angiogenesis and signalling by the pro-angiogenic growth factor VEGF (**Supplementary Figure 8**). We also applied DEPICT[19] with the full distribution of 5%FDR signals (**Online Methods**) to search for enriched gene sets. Blood vessel development, which includes angiogenesis, was in the top 10 ($p < 6.67x10^{-12}$ enrichment test[19]) DEPICT Grouped-GeneSets (GO:0001568; **Fig. 4, Supplementary Figure 9, Supplementary Table 11**).

Ingenuity built 5 networks out of the 357 genes with the largest three integrating 12 of the new candidate CAD risk genes with 67 candidate genes in known CAD loci (**Supplementary Table 12**). In total, the 5 networks comprise 66.4% of the 357 genes.

This is the largest CAD genetic study to assess simultaneously common and rare (MAF < 1%)/low-frequency (MAF 1-5%) variants. In total, 101 low-frequency and 3 rare variants reached genome-wide significance among all 5%FDR markers (N=11,427). This apparent paucity in rare variants which has also been reported for type 2 diabetes[20], is likely due to lack of power compared to studies of quantitative traits e.g. a study of adult height in ~700,000

individuals has reported 32 rare variants[5]. As expected, lower-frequency variants tend to have stronger effects compared to common variants (**Supplementary Figure 10**) with the exception of rs2891168 in *CDK2NB-AS1* (MAF 48.7%; OR 1.19; **Supplementary Table 13**). The intergenic variant rs186696265 which had the largest OR (1.62) in our study is known to affect LDL cholesterol levels[21].

Our findings highlight the importance of the FDR approach to define an extended list of associated variants. As we have previously proposed[1-2], suggestive association signals in well-powered GWAS such as this one can substantially improve our knowledge of disease architecture at only a modest penalty implied by the 5%FDR. We have demonstrated the potential value of the new 5%FDR list in improving prediction of CAD risk and implicating new networks underlying CAD pathophysiology. This extended list of candidate genes provides a powerful resource for functional studies.

We note that while this work was in review a study was published also reporting associations of the *HNF1A* locus with CAD[22].

URLs

www.ukbiobank.ac.uk/

GWAS catalogue: https://www.ebi.ac.uk/gwas/

GTEx portal: http://www.gtexportal.org/home/

PhenoScanner: http://www.phenoscanner.medschl.cam.ac.uk/

Ingenuity Knowledge Base: http://www.ingenuity.com/science/knowledge-base?utm_source=Blog&utm_medium=link&utm_campaign=Doug%20Bassett%20ASHG%202014

https://biobank.ctsu.ox.ac.uk/crystal/docs/genotyping_qc.pdf

http://cnsgenomics.com/shiny/INDI-V/

http://www.broadinstitute.org/mpg/depict

https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/

ACKNOWLEDGEMENTS

This work was funded by British Heart Foundation (BHF) grants RG/14/5/30893 to P.D. and FS/14/66/31293 to O.G. P.D. Work forms part of the research themes contributing to the translational research portfolios of the Barts Biomedical Research Center and Leicester Biomedical Research Centre funded by the National Institute for Health Research (NIHR). F.Y.L. and S.E.H. are funded by NIHR. C.P.N., T.R.W. N.J.S. are funded from BHF, Transatlantic Networks of Excellence Award (12CVD02) Leducq Foundation and EU-FP7/2007-2013 grant HEALTH-F2-2013-601456. N.J.S. is a NIHR Senior Investigator. PROCARDIS was supported by EU-FP6 (LSHM-CT- 2007-037273), AstraZeneca, BHF, Swedish Research Council, Knut and Alice Wallenberg Foundation, Swedish Heart-Lung Foundation, Torsten and Ragnar Söderberg Foundation, Karolinska Institutet, Foundation Strategic Research and Stockholm County Council (560283). M.F. and H.W. are supported by Wellcome Trust award 090532/Z/09/Z; M.F., H.W. and T.K., the BHF Centre of Research Excellence. A.G, H.W and T.K FP7/2007-2013 (HEALTH-F2-2013-601456 (CVGenes@Target), A.G. the Wellcome Trust and TriPartite Immunometabolism Consortium- Novo Nordisk Foundation's (NNF15CC0018486). HPS (ISRCTN48489393) was supported by: Medical Research Council (MRC), BHF, Merck and Co, Roche Vitamins Ltd. HPS acknowledges National Blood Service donor and the UK-Twin Study controls (WT 07611, FP7/2007-2013). J.C.H. is funded by BHF (FS/14/55/30806). Mount Sinai BioMe Biobank is supported by The Andrea and Charles Bronfman Philanthropies. GLACIER Study and P.W.F. are funded by European Commission (CoG-2015_681742_NASCENT), Swedish Research Council (Distinguished Young Researchers Award), Heart-Lung Foundation, and Novo Nordisk Foundation. OHGS studies were funded by Canadian Institutes of Health Research, Canada Foundation for Innovation and Heart & Stroke Foundation of Canada. LURIC was funded from the EU-FP7 (Atheroremo (201668), RiskyCAD (305739), INTERREG IV Oberrhein Program), European Regional Development Fund (ERDF), Wissenschaftsoffensive TMO, and from the German ministry for education and research, project e:AtheroSysMed (01ZX1313A-K). LOLIPOP is supported by NIHR-BRC Imperial College Healthcare NHS Trust, BHF (SP/04/002), MRC (G0601966, G0700931), WT (084723/Z/08/Z), NIHR (RP-PG-0407-10371), EU-FP7 (EpiMigrant, 279143) and Action on Hearing Loss (G51). The Helsinki Sudden Death Study was funded by EU-FP7 (201668, AtheroRemo), Tampere University Foundation, Tampere University Hospital Medical Funds (grants 9M048, 9N035 for T.L), the Emil Aaltonen Foundation (T.L), Finnish Foundation of Cardiovascular Research (T.L, P.K), Pirkanmaa Regional Fund of the Finnish Cultural Foundation, Yrjö Jahnsson Foundation, Tampere Tuberculosis Foundation (T.L), Signe and Ane Gyllenberg Foundation (T.L.), and Diabetes Research Foundation of Finnish Diabetes Association (T.L.). M.T. (PG/16/49/32176) and R.C. (FS/12/80/29821) are supported by BHF. E.Z. acknowledges WT funding (098051). H.S. was supported by Deutsche Forschungsgemeinschaft (Sonderforschungsbereich CRC 1123 (B02)). The MRC/BHF Cardiovascular Epidemiology Unit was funded by MRC (G0800270), BHF (SP/09/002), NIHR-BRC Cambridge, European Research Council (ERC; 268834), and EU-FP7 (HEALTH-F2-2012-279233), Pfizer, Merck and Biogen. EPIC-CVD was supported University of Cambridge, EU-FP7 (HEALTH-F2-2012-279233), MRC (G0800270) BHF (SP/09/002), and ERC (268834). We thank all EPIC participants and staff, Sarah Spackman for data management, and EPIC-CVD Coordinating Centre team.(IUT20-60) and Centre of Excellence in Genomics and Translational Medicine (GENTRANSMED), EU structural fund (Archimedes Foundation; 3.2.1001.11-0033), PerMed I and EU2020 (692145 ePerMed). This research was supported by

BHF (grant SP/13/2/30111) and conducted using the UK Biobank Resource (application number 9922).

AUTHOR CONTRIBUTIONS

Writing group (wrote and edited manuscript): C.P.N., A.G., A.S.B., S.K., T.R.W., E.M., I.N., J.C.H., O.G., H.S., M.F., J.D., N.J.S., H.W., P.D. All authors contributed and discussed the results, and commented on the manuscript. Data generation & cohorts: A.S.B., O.G., T.J., L.Z., S.E.H., E.A., T.L.A., E.P.B., J.C.C., R.C., R.M.C., P.E., R.E., E.E., P.W.F., C.G., D.G., A.H., J.M.M.H., E.I., A.K., T.Ke., T.Ky., T.L., X.L., Y.L., W.M., R.McP., A.M., C.N.A.P., M.Pujades-R., A.F.S., M.J.S., P.A.Z., K. A., R.J.F.L., E.Z., J.E., G.D., H.S., J.D., N.J.S., H.W., P.D. Phenotype data (UK Biobank, replication): C.P.N., A.S.B., I.N., F.Y.L., J.C.H., O.G., B.D.K., J.S.K., R.J.F.L., R.S.P., M.R., M.T., I.T., E.Z., J.E., G.D., H.S., J.D., N.J.S., H.W., P.D. Statistical analysis: C.P.N., A.G., A.S.B., S.K., T.J., M.F. Functional annotation: C.P.N., S.K., T.R.W., A.S.B., R.E., A.R., E.E.S., J.L.M.B. Biological and clinical enrichment and pathway analyses: E.M., P.D.

COMPETING FINANCIAL INTERESTS

P.W.F. has been a paid consultant for Eli Lilly and Sanofi Aventis and has received research support from several pharmaceutical companies as part of a European Union Innovative Medicines Initiative (IMI) project. E.I. is an advisor and consultant for Precision Wellness, Inc., and advisor for Cellink for work unrelated to the present project. M.K.R. has acted as a consultant for GSK, Roche, Ascensia, MSD, and also participated in advisory board meetings on their behalf. MKR has received lecture fees from MSD and grant support from Novo Nordisk, MSD and GSK. J.L.M.B. is the founder and chairman of Clinical Gene Networks. CGN has financially contributed to the STARNET study. J.L.M.B., E.E.S., and A.R. are on the board of directors for CGN. J.L.M.B. and A.R. own equity in CGN and receive financial compensation from CGN.

References

1. CARDIoGRAMplusC4D Consortium, Deloukas, P., Kanoni, S., Willenborg, C. *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45**, 25-33 (2013).

2. CARDIoGRAMplusC4D Consortium. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121-1130 (2015).

3. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N. Engl. J. Med*. **374**, 1134-1144 (2016)

4. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Systematic evaluation of pleiotropy identifies six further loci associated with coronary artery disease. *J. Am. Col. Cardiol*. **69**, 823-836 (2017)

5. Marouli, E. et al. Rare and low-frequency coding variants alter human adult height. *Nature* ***542***, 186-190 (2017)

6. Cardiovascular Disease Statistics 2015. British Heart Foundation

7. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369-375, S361-363 (2012)

8. Danesh, J., Saracci, R., Berglund, G., Feskens, E. *et al*. EPIC-Heart: the cardiovascular component of a prospective study of nutritional, lifestyle and biological factors in 520,000 middle-aged participants from 10 European countries. *Eur. J. Epidemiol.* **22**, 129-141 (2007)

9. Wipff, P.J., Rifkin, D.B., Meister, J.J., Hinz, B. Myofibroblast contraction activates latent TGF-beta1 from the extracellular matrix. *J. Cell Biol.* **179**, 1311-23(2007)

10. Henderson, N.C., Arnold, T.D., Katamura, Y., Giacomini, M.M. *et al*. Targeting of αv integrin identifies a core molecular pathway that regulates fibrosis in several organs. *Nat. Med.* **19**, 1617-24 (2013)

11. Goumans, M.J., Liu, Z., ten Dijke, P. TGF-beta signaling in vascular biology and dysfunction. *Cell Res.* **19**, 116-27 (2009)

12. Nanda, V., Downing, K.P., Ye, J., Xiao, S. *et al*. CDKN2B Regulates TGFβ Signaling and Smooth Muscle Cell Investment of Hypoxic Neovessels. *Circ. Res.* **118**, 230-40 (2016).

13. Franzén, O., Ermel, R., Cohain, A., et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827-830 (2016)

14. Kukreja, R.C., Salloum, F.N., Das, A. Cyclic guanosine monophosphate signaling and phosphodiesterase-5 inhibitors in cardioprotection. *J. Am. Coll. Cardiol.* **59**, 1921–1927 (2012)

15. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multi-tissue gene regulation in humans. *Science* **348**, 648-660

16. Samson, T., van Buul, J.D., Kroon, J., Welch C, *et al*. The guanine-nucleotide exchange factor SGEF plays a crucial role in the formation of atherosclerosis. *PLoS One* **8**, e55202 (2013)

17. Babaev, V.R., Porro, F., Linton, M.F., Fazio, S. *et al*. Absence of regulated splicing of fibronectin EDA exon reduces atherosclerosis in mice. *Atherosclerosis* **197**, 534-540 (2008)

18. Rohwedder, I., Montanez, E., Beckmann, K., Bengtsson, E. *et al*. Plasma fibronectin deficiency impedes atherosclerosis progression and fibrous cap formation. EMBO Mol. Med. **4**, 564-576 (2012)

19. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).

20. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-47 (2016)

21. Surakka, I. *et al*. The impact of low-frequency and rare variants on lipid levels. *Nat. Genet.* **47**, 589-97 (2015)

22. Howson, J. M. M. *et al*. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat. Genet.* (2017)

Figure legends

Figure 1. Description of HARD and SOFT CAD phenotypes in UK Biobank. (**a**) Diagram depicting the CAD phenotype definition in UK Biobank. HARD CAD defined as fatal or non-fatal myocardial infarction (MI), PTCA (percutaneous transluminal coronary angioplasty), or coronary artery bypass grafting (CABG). SOFT CAD includes HARD CAD as well as chronic ischaemic heart disease (IHD) and angina. UK Biobank self-reported data: 'Vascular/heart problems diagnosed by doctor' or 'Non-cancer illnesses that self-reported as angina or heart attack'. Self-reported surgery defined as either PTCA, CABG or triple heart bypass. HESIN hospital episodes data and death registry data using diagnosis and operation - primary and secondary cause: MI defined as hospital admission or cause of death due to ICD9 410-412, ICD10 I21-I24, I25.2; PTCA is defined as hospital admission for PTCA (OPCS-4 K49, K50.1, K75); CABG is defined as hospital admission for CABG (OPCS-4 K40 – K46); Angina or chronic IHD defined as hospital admission or death due to ICD9 413, 414.0, 414.8, 414.9, ICD10 I20, I25.1, I25.5-I25.9. (**b**) Radar plot highlighting the proportions (%) of signals between the HARD and SOFT CAD phenotype definitions based on the 5%FDR results (**Supplementary Table 4)**; MAF = minor allele frequency, $p < 5x10^{-8}$ marks variants reaching genome-wide significance, OR = odds ratio (OR > 1.05 corresponds to 85% power to detect a signal (alpha < 0.05) in the SOFT analysis). The results for all six subgroups of variants assessed did not differ statistically between the two phenotype definitions (p>0.1)

Figure 2. Transposed Manhattan plot showing the SOFT meta-analysis results under an additive model. The *P*-values are truncated at –log10(*P*) = 20. Markers shown are from the meta-analysis of UK Biobank with the 1000G GWAS data[2] unless flagged by an * (exome chip markers). The red dotted lines are at GWAS (P=5x10^{-8}) and 5% FDR significance (P=6.28x10^{-5}). The known CAD risk loci are shown in black (**Supplementary Table 2**); *KSR2* and *ZNF507*-

435 *LOC400684* had reached genome-wide significance under a recessive model[2]. The
436 11p15_MRVI1 / CTR9 locus had discordant results between the CAD 1000 Genomes GWAS[2]
437 and Exome[4] data set. The lead variant in the Exome data set, rs11042937, had $P = 3.21 \times 10^{-8}$;
438 data shown are from the meta-analysis with the 1000Genomes GWAS as this marker had
439 an imputation info score of 1 (Online Methods). The 13 novel CAD loci which reached genome-
440 wide significance in our study (including replication data; **Table 1**), are written in brown font.
441 **Figure 3.** Single marker p-value comparison of the 5% FDR variants in the published
442 CARDIoGRAMplusC4D 1000Genomes CAD GWAS meta-analysis[2] and current FDR study. Of
443 the 162 variants which had $p < 5 \times 10^{-5}$ in the CAD 1000Genomes GWAS, 116 had a match or
444 good proxy ($r^2 > 0.8$) in the new FDR list (blue circles). SNPs in red (n=7) were present in the
445 earlier FDR list and reached genome-wide significance in the current analysis.
446 **Figure 4**. Heat map showing the DEPICT gene set enrichment results with zoom-in on a subset
447 of the results. 556 gene sets are included which had evidence of enrichment at 1% FDR. The
448 x– axis shows the gene name, which is predicted to be included in the reconstituted gene set
449 indicated in the y – axis. The color red indicates higher Z-score, where Z-score is a value
450 representing each gene's inclusion in DEPICT's reconstituted gene sets. Clustering was made
451 based on complete linkage method. Highlighted pathways in the cluster, include
452 angiogenesis, blood vessel development and morphogenesis.
453
454

Table 1-Novel variants reaching genome-wide significance ($P<5x10^{-8}$) in the combined (discovery and replication) SOFT meta-analysis

Locus Name	Markername	CHR	POS (hg38)	EA	EAF	Functional Evidence	UKBB+CoG/Exome				Meta analysis		
							OR	95% CI	Pvalue	FDR Qvalue	OR	95% CI	Pvalue
TDRKH	rs11810571	1	151762308	G	0.849	eQTL/coding	1.060	1.039-1.082	$2.21x10^{-8}$	$8.05x10^{-5}$	1.057	1.036-1.079	$4.24x10^{-8}$
FN1	rs1250229*	2	216304384	T	0.256	eQTL/coding	1.072	1.052-1.092	$1.85x10^{-13}$	$2.05x10^{-9}$	1.071	1.051-1.091	$2.77x10^{-13}$
RHOA	rs7623687	3	49448566	A	0.855	none	1.074	1.049-1.100	$3.72x10^{-9}$	$1.62x10^{-5}$	1.076	1.052-1.101	$3.44x10^{-10}$
UMPS/ITGB5	rs142695226	3	124475201	G	0.138	eQTL/coding	1.069	1.045-1.094	$1.00x10^{-8}$	$3.98x10^{-5}$	1.071	1.048-1.095	$1.53x10^{-9}$
ARHGEF26	rs12493885*	3	153839866	C	0.886	eQTL	1.074	1.047-1.101	$3.29x10^{-8}$	$1.15x10^{-4}$	1.073	1.047-1.101	$3.16x10^{-8}$
PRDM8/FGF5	rs10857147	4	81181072	T	0.275	none	1.056	1.036-1.075	$8.96x10^{-9}$	$3.60x10^{-5}$	1.054	1.036-1.073	$5.66x10^{-9}$
PDE5A/MAD2L1	rs7678555	4	120909501	C	0.301	eQTL	1.049	1.031-1.069	$1.43x10^{-7}$	$4.25x10^{-4}$	1.052	1.034-1.070	$1.32x10^{-8}$
HDGFL1	rs6909752	6	22612629	A	0.351	none	1.051	1.034-1.069	$5.59x10^{-9}$	$2.35x10^{-5}$	1.051	1.034-1.068	$2.19x10^{-9}$
ARNTL	rs3993105	11	13303071	T	0.704	none	1.048	1.030-1.067	$1.06x10^{-7}$	$3.33x10^{-4}$	1.048	1.031-1.066	$4.77x10^{-8}$
HNF1A	rs2244608	12	121416988	G	0.355	coding	1.053	1.035-1.070	$2.32x10^{-9}$	$1.06x10^{-5}$	1.053	1.035-1.070	$7.74x10^{-10}$
CDH13	rs7500448	16	83045790	A	0.752	eQTL	1.061	1.040-1.082	$5.14x10^{-9}$	$2.18x10^{-5}$	1.063	1.043-1.083	$4.76x10^{-10}$
TGFB1	rs8108632	19	41854534	T	0.488	none	1.049	1.031-1.067	$5.88x10^{-8}$	$1.95x10^{-4}$	1.048	1.031-1.066	$4.04x10^{-8}$
SNRPD2	rs1964272	19	46190268	G	0.510	none	1.045	1.028-1.063	$2.29x10^{-7}$	$6.15x10^{-4}$	1.047	1.030-1.064	$2.46x10^{-8}$

*Exome marker

EA: effect allele; EAF: Effect allele frequency; CoG = CARDIoGRAMplusC4D 1000G GWAS; Exome = Exome array analysis; UKBB = UK Biobank; Discovery sample comprised 71,602 cases and 260,875 controls (for exome markers 53,135 and 215,611 respectively); Replication sample comprised up to 4412 cases and 3910 controls. Functional evidence for the locus is given where the lead variant or a variant in high LD ($r^2>0.8$) is a coding change, has evidence as an expression quantitative trait locus (eQTL), or both. Further details of functional evidence are provided in **Supplementary Table 7** and **Supplementary Figure 6**.

Online Methods

Phenotype Definitions & Power calculation

UKBB recruited 502,713 individuals aged 40-69 years from England, Scotland and Wales between 2006 and 2010 (94% of self-reported European ancestry). HARD CAD was defined as fatal or non-fatal myocardial infarction (MI), percutaneous transluminal coronary angioplasty (PTCA), or coronary artery bypass grafting (CABG). SOFT CAD includes all HARD CAD as well as chronic ischemic heart disease (IHD) and angina. Controls were defined as patients which were not a SOFT case after exclusions (listed below). All conditions were defined by either self-reported, hospital episode or death registry data.

Exclusions were made for aneurysm and atherosclerotic cardiovascular disease using hospital admissions, or cause of death, codes ICD9 414.1, ICD 10 I25.0, I25.3, I25.4, and not having MI, PTCA, CABG, Angina or chronic IHD as defined above.

Susceptibility effect sizes in MI cases and an inclusive CAD definition were very similar in the earlier GWAS[2]. We hypothesized that the detailed clinical information in UKBB might enhance the search for novel loci by further broadening the CAD phenotype to increase sample size.

GWAS and meta-analyses

All participants gave written consent for participation in genetic studies, and the protocol of each study was approved by the corresponding local research ethics committee or institutional review board. Participating cohorts in the 1000 Genomes and Exome GWAS studies are described elsewhere[2,3]. UK Biobank (UKBB samples) were excluded due to withdrawn consent, sex mismatches (n=182), Biobank/Believe QC exclusions (n=406) and sample relatedness (n=3,481) determined as Kinship>0.088. GWAS analysis in UKBB was restricted to variants with results available in the published GWAS[2] or Exome[3-4] dataset. Further exclusions included poorly imputed (info<0.4) or monomorphic variants, duplicate variants across data sets, variants that deviated strongly from Hardy-Weinberg Equilibrium in European ancestry controls ($p<1x10^{-9}$), variants with an effect allele frequency in European ancestry samples that differed strongly (i) from 1000G European panel, (ii) from GWAS/Exome data, (iii) between arrays (UKBB vs UK-BiLEVE), and (iv) across genotyping batches. Variants that did not produce a valid result or estimated extreme log odds ratios (|beta|>4) were also excluded after analysis. Cluster plots lead variants and of proxies were visually inspected.

We ran the GWAS under an additive frequentist mode of inheritance for each variant using the dosages from the imputed data, adjusting for array (UK Biobank vs UK BiLEVE) and the first five principal components (see **URLs**) using SNPTEST. Age and sex were not adjusted for to maximize the power to detect associations with diseases that have a prevalence <10%[23]. Population stratification was assessed and standard errors were adjusted using the genomic inflation statistic (λ).

Association summary statistics (after λ correction) from the UKBB were combined with the 1000 Genomes (1000G) imputed GWAS results[2] and the Exome results[3] via two separate fixed-effect inverse-variance weighted meta-analysis implemented in GWAMA[24]. We applied post meta-analysis λ correction in each instance. We identified 36,460 variants present in both the 1000G imputed GWAS and the Exome results. We retained the variants from the 1000G imputed GWAS if the median info score was 1, otherwise we retained the results from the Exome data.

Comparison of SOFT vs HARD peak variant lists at 5% q-value

The false discovery rate (FDR) following the meta-analysis with UKBB was assessed using a step-up procedure in the *qqvalue* Stata program[25] as it is well controlled under positive regression-dependency conditions. We used the Simes method to generate q-values for the 8.9M variants. The p-value cut-off for a q-value of 5% for HARD was 7.24×10^{-5} and SOFT was 6.28×10^{-5}. Peak SNPs were identified in a 1cM window. There is an exact overlap of 155 variants between the 2 peak variant lists, however, using the 1cM window the overlap increases to 206 variants. Both the lists were annotated and classified into 6 categories (exome chip, indels, Odds Ratio (OR)>1.05, p<5e-8, MAF<5% and exonic). The proportions were calculated in each of the 6 categories and plotted as a radar plot (**Fig. 1b**). Monte Carlo simulations were used to assess the *post-hoc* power of the UKBB interim data to replicate the 155 variants. The 1000G GWAS effect sizes ("betas") are expected to be subject to *winner's curse* inflation so were shrunken (towards the null) by application of the FIQT procedure[26]. Effect sizes for firmly established CAD loci were systematically lower for SOFT compared to the HARD phenotype (**Supplementary Table 1**) noting that HARD closely corresponds to the CAD phenotype in reference 2. Betas were therefore further shrunken by a factor log (1.059)/log(1.072) = 0.82 (**Supplementary Table 1**). 10,000 replicates were then randomly drawn from the vector of shrunken betas and the corresponding UKBB standard errors, to allow for variation in genotype call rates, imputation quality and allele frequency and to calculate Wald association statistics. Multiple testing of 155 variants was allowed for by controlling the FDR to 5% with a step-up procedure encoded in the *multproc*[27] Stata™ program. The average expected number of replicated variants was 56 (95%CI 42 – 69). Testing the 5% FDR variants (**Supplementary Table 7**) in UKBB with a model adjusted for age and sex gave concordant results to the unadjusted model (data not shown).

GCTA & Heritability analysis

We used the GCTA software[7] to perform joint association analysis in (SOFT) meta-analysis results. This approach fits an approximate multiple regression model using summary-level meta-analysis statistics and LD corrections estimated from a reference panel (here the UKBB sample). We adopted a chromosome-wide stepwise selection procedure to select variants and estimate their joint effects at i) a genome-wide significance level (pJoint ≤ 5×10^{-8}) in the totality of meta-analysed variants (n~ 9M; **Supplementary Figure 10**, **Supplementary Table 11**) and ii) a Bonferroni-corrected pJoint<1×10^{-4} corresponding to the number of independent LD bins ($r^2 < 0.1$) in the 5% FDR variant list (n=11,427; **Supplementary Table 6**).

Heritability calculations were based on a multifactorial liability-threshold model, implemented in the INDI-V[28] calculator (see **URLs**), under the assumption of a baseline population risk (K) of 0.0719[29] and a twins heritability (H_L^2) of 0.4. Multiple regression estimates from the GCTA joint association analysis were used to estimate heritability for the 304 independent CAD risk variants within the 5% FDR list.

Genetic risk score analysis

GRS analysis was undertaken in the EPIC-CVD study[8] which comprises 7910 CAD cases and 12958 controls (**Supplementary Note**). We considered either all known and new lead CAD

risk variants reaching genome-wide significance (GRS2; **Supplementary Table 2** and **Table 1**) or the 304 variants in the 5% FDR set (GRS1; **Supplementary Table 7**). We used variants with an INFO score filter of 0.4 in EPIC-CVD and replaced missing ones with proxies (r2 > 0.8 in 1000 Genomes European participants). GRS1 comprised 280 variants and GRS2 71. The raw GRS was obtained by summing the dosages of these variants for all individuals. We then fitted a Prentice weighted cox regression model for each GRS, adjusting for age and sex, to obtain survival forecasts and calculate the C indices. Statistical analyses were performed using R 3.3.3 and STATA 13.1. Variant extraction was done using qctool 1.4.

Functional annotation

eQTLs: For associations between the 304 independent variants (5% FDR) and gene expression traits we searched for expression quantitative trait loci (eQTLs) in the Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET) RNA-seq dataset[13] and the Genotype-Tissue Expression[15] (GTEx) portal. eQTLs were included if the best eSNP (i.e. the variant with the most significant association with gene expression in cis) was in high LD (r^2>0.8) with the CAD lead SNP.

Regulatory elements: We functionally annotated each of the 13 lead variants and their proxies (r^2>0.8) using HaploregV4[30]. Overlap with regulatory elements including chromosome state segmentation, DNase hypersensitivity, and transcription factor binding (TFB) as determined by the ENCODE[31] and Roadmap Epigenome projects[32], and predicted effects on TFB based on regulatory motifs from TRANSFAC[33] and JASPAR[34] were identified using HaploregV4[19] and the UCSC genome browser. Variants were then scored using three different bioinformatics tools that help prioritise causal disease variants. Combined Annotation Dependent Depletion (CADD)[35] incorporates a range pathogenicity prediction tools to provide a genome-wide score (C-score) for each test variant from its pre-calculated database of ~8.6 billion genetic variants. High scores indicate variants that are not stabilized by selection and are more likely to be disease-causing and low scores indicate evolutionary stable non-damaging variants. The top 10% of likely functional variants will have a C-score >10 and top 1% of variants will have a C-score >20. Genome-wide annotation of variants (GWAVA)[36] predicts the functional impact of noncoding variants based on genomic and epigenomic annotations and provides scores between 0 and 1 with higher scores indicating variants that are more likely to be functional. RegulomeDB[37] annotates and scores variants in seven categories based datasets such as ENCODE. Scores of 1-2 variants likely to affect TFB, 3 less likely to affect binding, 4-6 relate to variants with minimal binding evidence and 7 is for variants with no regulatory annotation.

Phenome-scan: look ups in other common traits were performed using the PhenoScanner database as described in reference 38.

Pathway analysis

DEPICT: DEPICT[19] is a computational tool which performs gene set enrichment analyses to prioritize genes in associated GWAS loci with probabilistically predefined gene sets based on Gene Ontology terms, canonical pathways, protein-protein interaction subnetworks and rodent phenotypes; reconstituted gene sets are detailed in references 19 and 39. Input to our analysis were the 11,427 CAD variants (FDR 5%) of which 11,311 were annotated in DEPICT. We constructed loci as previously described (beta version 1.1, release 194, see **URLs**). Analysis was performed with default parameters (50 repetitions to compute FDRs, 500 permutations to adjust for biases, such as gene length). The 11,311 variants were

collapsed to 288 loci which were used in the gene set enrichment analyses. Correlated gene sets were grouped together based on gene membership to expedite data interpretation.

Ingenuity: Genes were selected using 304 independent SNPs (5% FDR) based on eQTLs (**Supplementary Table 9**) and physical proximity (included overlapping genes on opposite strands or at equal distance from the SNP). Spliced ESTs and putative transcripts were not included. Network analysis was performed using the Ingenuity Pathway Analysis software (see **URLs**). We considered molecules and or relationships available in The IPA Knowledge Base (IKB) for human OR mouse OR rat and set the confidence filter to Experimentally Observed OR High (Predicted). Networks were generated with a maximum size of 70 genes and up to 10 networks were allowed. Networks are ranked according to their degree of relevance to the 'eligible' molecules in the query data set. The network score is based on the hypergeometric distribution and is calculated with the right-tailed Fisher's Exact Test. The significance p-value associated with enrichment of functional processes is calculated using the right-tailed Fisher Exact Test by considering the number of query molecules that participate in that function and the total number of molecules that are known to be associated with that function in the IKB.

Data Availability Statement: Meta-analysis summary statistics for all variants considered in this study for association with CAD (SOFT definition) are available at http://www.cardiogramplusc4d.org/data-downloads/.

Method References

23. Pirinen, M., Donnelly, P. & Spencer, CC. Including known covariates can reduce power to detect genetic effects in case-control studies. *Nat Genet* **44**, 848-51 (2012).

24. Magi, R. & Morris, AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*, **11**:288 (2010).

25. Newson, R. QQVALUE: Stata module to generate quasi-q-values by inverting multiple-test procedures, Statistical Software Components S457100, Boston College Department of Economics, (2009, revised 2013)

26. Bigdeli, T.B. *et al*. A simple yet accurate correction for winner's curse can predict signals discovered in much larger genome scans. *Bioinformatics* **32**, 2598-603 (2016)

27. Newson R. Multiple-test procedures and smile plots. *The Stata Journal* **3**, 109-132 (2003).

28. Witte, J.S., Visscher, P.M. & Wray, N.R. The contribution of genetic variants to disease depends on the ruler. *Nat Rev Genet* **15,** 765-76 (2014).

29. Muñoz, M. *et al.* Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat Genet* **48**, 980-3 (2016).

30. Ward, L.D., Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 40, D930-4 (2012).

31. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

32. Skipper, M., Eccleston, A., Gray, N., *et al*. Presenting the epigenome roadmap. *Nature* **518**, 313 (2015).

33. Matys, V., Fricke, E., Geffers, R., *et al*. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31:374–8 (2003).

34. Portales-Casamar, E., Thongjuea, S., Kwon, A.T., *et al*. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **38**, D105–10 (2010).

35. Kircher, M., Witten, D.M., Jain, P., *et al*. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).

36. Ritchie, G.R., Dunham, I., Zeggini, E., *et al*. Functional annotation of noncoding sequence variants. *Nat Methods* **11**, 294-6 (2014).

37. Boyle, A.P., Hong, E.L., Hariharan, M., et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research* **22**, 1790-7 (2012)

38. Staley, J.R. *et al*. PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016)

39. Fehrmann, R. S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* **47,** 115-125 (2015)

668