



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/116865/>

Version: Accepted Version

Article:

Nelson, CP, Goel, A, Butterworth, A et al. (2017) Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature Genetics*, 49 (9). pp. 1385-1391. ISSN: 1061-4036

<https://doi.org/10.1038/ng.3913>

© 2017 Nature America, Inc., part of Springer Nature. This is an author produced version of a paper published in *Nature Genetics*. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

1 Association analyses based on false discovery rate implicate 243 2 susceptibility loci for coronary artery disease

3 Christopher P. Nelson^{1,2,61}, Anuj Goel^{3,4,61}, Adam Butterworth^{5,6,61}, Stavroula Kanoni^{7,8,61}, Tom
4 R. Webb^{1,2}, Eirini Marouli^{7,8}, Lingyao Zeng⁹, Ioanna Ntalla^{7,8}, Florence Y. Lai^{1,2}, Jemma C.
5 Hopewell¹⁰, Olga Giannakopoulou^{7,8}, Tao Jiang⁵, Stephen E. Hamby^{1,2}, Emanuele Di
6 Angelantonio^{5,6}, Themistocles L. Assimes¹¹, Erwin P. Bottinger¹², John C. Chambers^{13,14,15},
7 Robert Clarke¹⁰, Colin Palmer^{16,17}, Richard M. Cubbon¹⁸, Patrick Ellinor¹⁹, Raili Ermel²⁰,
8 Evangelos Evangelou^{21,22}, Paul W. Franks^{23,24,25}, Christopher Grace^{3,4}, Dongfeng Gu²⁶, Aroon
9 D. Hingorani²⁷, Joanna M. M. Howson⁵, Erik Ingelsson²⁸, Adnan Kastrati⁹, Thorsten Kessler⁹,
10 Theodosios Kyriakou^{3,4}, Terho Lehtimäki²⁹, Xiangfeng Lu²⁶, Yingchang Lu^{12,30}, Winfried
11 März^{31,32,33}, Ruth McPherson³⁴, Andres Metspalu³⁵, Mar Pujades-Rodriguez³⁶, Arno
12 Ruusalepp^{20,37}, Eric E. Schadt³⁸, Amand F. Schmidt³⁹, Michael J. Sweeting⁵, Pierre A.
13 Zalloua^{40,41}, Kamal AlGhalayini⁴², Bernard D. Keavney^{43,44}, Jaspal S. Kooner^{14,15,45}, Ruth J. F.
14 Loos^{12,46}, Riyaz S. Patel^{47,48}, Martin K. Rutter^{49,50}, Maciej Tomaszewski^{51,52}, Ioanna Tzoulaki^{21,22},
15 Eleftheria Zeggini⁵³, Jeanette Erdmann^{54,55,56}, George Dedoussis⁵⁷, Johan L. M.
16 Björkegren^{37,38,58}, EPIC-CVD Consortium, CARDIoGRAMplusC4D, The UK Biobank
17 CardioMetabolic Consortium CHD working group, Heribert Schunkert^{9,61}, Martin Farrall^{3,4,61},
18 John Danesh^{5,6,59,61}, Nilesh J. Samani^{1,2,61}, Hugh Watkins^{3,4,61}, Panos Deloukas^{7,8,60,61}

- 19 1. Department of Cardiovascular Sciences, University of Leicester, Leicester LE3 9QP, UK
20 2. National Institute for Health Research Leicester Cardiovascular Biomedical Research Unit, Leicester LE3 9QP,
21 UK
22 3. Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3
23 9DU, UK
24 4. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK
25 5. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of
26 Cambridge, Cambridge CB1 8RN, UK
27 6. NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and
28 Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
29 7. William Harvey Research Institute, Barts & the London Medical School, Queen Mary University of London,
30 London EC1M 6BQ, UK
31 8. Centre for Genomic Health, Queen Mary University of London, London EC1M 6BQ, UK
32 9. German Heart Center Munich, Clinic at Technische Universität München, Munich 80636, Germany
33 10. CTSU, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK
34 11. Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA
35 12. Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York
36 City, NY 10029, USA
37 13. Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK
38 14. Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, Middlesex UB1 3HW,
39 UK
40 15. Imperial College Healthcare NHS Trust, London W12 0HS, UK
41 16. Molecular and Clinical Medicine, Biomedical Research Institute, University of Dundee, Ninewells Hospital,
42 Dundee, UK
43 17. Pharmacogenomics Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee
44 18. Leeds Institute of Cardiovascular and Metabolic Medicine, The University of Leeds, Leeds LS2 9JT, UK
45 19. Cardiac Arrhythmia Service and Cardiovascular Research Center, The Broad Institute of Harvard and MIT,
46 Boston, USA
47 20. Department of Cardiac Surgery, Tartu University Hospital, Tartu 50406, Estonia
48 21. Department of Biostatistics and Epidemiology, Imperial College London, London W2 1PG, UK
49 22. Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina 45110,
50 Greece
51 23. Department of Clinical Sciences, Genetic & Molecular Epidemiology Unit, Lund University Diabetes Center,
52 Skåne University Hospital, Lund University, Malmö/Skåne SE-205 02, Sweden
53 24. Department of Nutrition, Harvard T. H. Chan School of Public Health, Harvard University,
54 Boston/Massachusetts MA 02115, USA

- 55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
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 School of Medicine at Mount Sinai, New York, NY 10069, USA, New York 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. Cardiovascular, Metabolic and Nutritional Sciences Research Domain/University of Manchester, Manchester M13 9PL, UK
 50. Manchester Diabetes Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 0JE, England
 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

111

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

113 **Genome-wide association studies (GWAS) in coronary artery disease (CAD) have identified**
114 **66 loci at 'genome-wide significance' ($p < 5 \times 10^{-8}$) but a much larger number of putative loci**
115 **at a false discovery rate (FDR) of 5%¹⁻⁴. Here, we leverage an interim release of UK Biobank**
116 **(UKBB) data to evaluate the validity of the FDR approach. We tested a CAD phenotype**
117 **inclusive of angina (SOFT; $N_{\text{cases}}=10,801$) as well as a stricter definition without it (HARD;**
118 **$N_{\text{cases}}=6,482$) and selected the former for conducting a meta-analysis with the two most**
119 **recent CAD GWASs²⁻³. This approach identified 13 new loci at genome-wide significance, 12**
120 **of which were in our previous 5% FDR list², and provided strong support that the remaining**
121 **FDR loci represent genuine signals. The set of 304 independent variants at 5% FDR in this**
122 **study explain 21.2% of CAD heritability and identified 243 loci that implicate pathways in**
123 **blood vessel morphogenesis as well as lipid metabolism, nitric oxide signaling and**
124 **inflammation.**

125

126

127 Previous GWAS studies of CAD risk¹⁻⁴ have interrogated a large number of cases and controls
128 but remain less well-powered than GWAS of quantitative traits⁵. UKBB was established to
129 improve understanding of the causes of common diseases including CAD, a leading health
130 problem around the world⁶. In addition to self-reported disease outcomes and extensive
131 health and life-style questionnaire data, the 502,713 participants are being tracked through
132 their NHS records and national registries (including cause of death and Hospital Episode
133 Statistics). In July 2015, UKBB released genotypes imputed to the 1000 Genomes panel for
134 152,249 participants profiled with a SNP array harboring 820,967 variants comprising
135 common variants optimized for imputation, validated rare coding variants and sets of
136 phenotype-associated variants or their proxies (e.g. GWAS catalogue).

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

140 The July 2015 release of UKBB comprises 10,801 genotyped individuals with an inclusive CAD
141 phenotype ('SOFT') that incorporates self-reported angina or other evidence of chronic
142 coronary heart disease, of which 6,482 have a more stringently defined CAD phenotype
143 ('HARD') of myocardial infarction and/or revascularisation (**Fig. 1a**). After QC we analysed the
144 SOFT and HARD cases separately against 137,914 controls for 9,149,595 variants present
145 either in the CARDIoGRAMplusC4D 1000-Genomes GWAS² or the MIGen/CARDIoGRAM
146 Exome-chip study³⁻⁴. The SOFT definition was selected for the primary analysis based on
147 power calculations (**Supplementary Table 1**). We found 4 (SOFT and HARD), 1 (SOFT only) and
148 2 (HARD only) variants reaching genome-wide significance, all located in known CAD loci
149 (**Supplementary Figure 1**).

150 We then meta-analysed the UKBB data for each CAD definition with each of the two published
151 data sets (**Supplementary Figure 2**) applying a double genomic control correction. For both
152 the SOFT and HARD definitions, we validated all 66 known CAD loci (72 independent variants
153 with $p < 1.2 \times 10^{-3}$) with 43 and 37 respectively reaching genome-wide significance in this
154 study (**Supplementary Table 2**). Outside the known CAD loci (1 Mb window centred on the
155 published lead SNP) we found 9 new signals (in both SOFT and HARD) reaching genome-wide
156 significance (**Table 1** and **Fig. 2**). The anticipated increase in power with the SOFT definition
157 (**Supplementary Table 1**) was attenuated by an inflation of the lambda statistic
158 (**Supplementary Table 3**), potentially due to a combination of larger sample size (i.e.
159 polygenicity) and a less homogeneous phenotype in the SOFT definition. Overall, there was
160 strong concordance between corresponding signals for SOFT and HARD (**Fig. 1b**,

161 **Supplementary Table 4**); subsequent analyses were undertaken using the SOFT meta-analysis
162 results.

163 To look for additional signals beyond the 9 that reached genome-wide significance (**Fig. 2**) we
164 performed an FDR analysis and selected 23 suggestive signals at 1% FDR ($p < 1.55 \times 10^{-6}$;
165 **Supplementary Table 4**) outside known CAD loci which we validated in an independent
166 sample of up to 4,412 cases and 3,910 controls from the German MI-Family-Studies V and VI
167 and a Greek case-control study (**Supplementary Table 5**). In total, we identified 13 new
168 genome-wide significant CAD loci in the combined discovery and replication sample (**Table 1**,
169 **Supplementary Table 6**).

170 In our recent large-scale GWAS², we reported 162, mainly common, variants at an FDR
171 discovery cutoff of 5% showing conditional independent associations with the P_{joint} test in
172 GCTA⁷. Twelve of the 13 new sentinel SNPs were present or had a proxy ($r^2 > 0.8$) among these
173 162 variants². **Fig. 3** shows a strong linear relationship between association signals for these
174 162 variants in the earlier² and current analysis, with overall greater significance levels in the
175 current meta-analysis. As expected, we observed an excess of small p-values for this set of
176 variants in the UK Biobank alone (**Supplementary Figure 3a**). Monte Carlo simulations show
177 that the expected number of replicated variants in the UK Biobank data is 56 (95%CI 42 – 69)
178 (**Supplementary Figure 3b**) and we found 58 variants after allowing for multiple testing (q-
179 values < 0.05). This further confirms the validity of extended lists of associated variants based
180 on FDR criteria. We therefore defined a new FDR list of association signals by performing an
181 approximate joint association analysis with the GCTA software⁷ as described elsewhere² using
182 the 11,427 SNPs with 5%FDR. We identified 304 independent variants at $P_{\text{joint}} < 10^{-4}$, clustering
183 in 243 putative CAD loci (**Supplementary Table 7**). The new 5%FDR set overlaps by 122 SNPs
184 with the old set (75.3%; including proxies at an $r^2 > 0.8$). We then assessed heritability using

185 the independent set of 304 SNPs and obtained a heritability estimate of 21.2%. The
186 contribution to this heritability estimate of the 13 new loci (**Table 1**) was 1.03% whereas the
187 new and known genome-wide significant CAD loci together explained 8.53% of CAD
188 heritability. To further assess the validity and utility of the 5%FDR set, we tested the ability to
189 predict CAD using genetic risk scores (GRS) based on either the 5%FDR SNPs (GRS1) or only
190 CAD variants reaching genome-wide significance (GRS2; **Online Methods**) in an independent
191 sample, EPIC-CVD⁸, comprising 7910 CHD cases and 12958 controls. In a model with age and
192 sex, GRS1 increased the C-index by 0.25% compared to GRS2 (**Supplementary Table 8**). GRS1
193 improved the point estimates of the HR compared to GRS2 mainly in the second (from 0.9116
194 to 0.8314) and fourth quintile (from 1.0437 to 1.176), **Supplementary Figure 4**.

195 We then explored the biology of the 13 new genome-wide significant CAD risk loci;
196 **Supplementary Figure 5** shows regional association plots. **Supplementary Figure 6** provides
197 *in silico* functional annotation (**Online Methods**) for each lead variant and its proxies (1000
198 Genomes). We found compelling evidence to implicate candidate genes *ITGB5*, *TGF1*, *PDE5A*,
199 *ARHGEF26*, *FN1*, *CDH13*, and *HNF1* (detailed in **Supplementary Note**). The risk allele of
200 rs150512726 (proxy for rs142695226; **Table 1**), causes a 3 amino acid deletion within the
201 cytoplasmic tail of integrin subunit beta 5 (ITGB5), part of a heterodimer which regulates the
202 activation of latent TGF1 (Transforming growth factor beta 1)⁹⁻¹⁰. The intronic variant
203 (rs8108632; **Table 1**) we identified in *TGF1*, further implicates the TGF1 pathway in CAD
204 risk. TGF1 is known to have important roles in endothelium and vascular smooth muscle¹¹
205 but has not been widely studied in atherosclerosis, though a recent study implicates TGF1
206 signalling downstream of CDKN2B in the *CDKN2BAS* cardiovascular risk locus¹². eQTL analyses
207 suggested candidate CAD risk genes (*TDRKH*, *FN1*, *ARHGEF26*, *PDE5A*, *ARNTL*, and *CDH13*) in
208 six new loci (**Supplementary Table 9**). For example, the lead variant rs7678555 (**Table 1**) was

209 found to be a strong eQTL ($p=8.1 \times 10^{-13}$) for *PDE5A* only in aorta from CAD patients
210 (STARNET¹³; **Supplementary Table 9**) although its regulatory potential was modest using
211 functional prediction tools (**Online methods**). *PDE5A* encodes a cGMP-specific
212 phosphodiesterase which is important for smooth muscle relaxation in the cardiovascular
213 system where it regulates nitric-oxide-generated cGMP¹⁴. Furthermore, mining eQTL data in
214 tissues from CAD patients (STARNET) showed several other instances of eSNPs (*TDRKH*, *FN1*,
215 *CDH13*; **Supplementary Table 9**) having no effect in tissues from non-CAD patients (GTEx¹⁵),
216 highlighting the need to expand efforts to map regulatory elements in disease tissues.
217 Other candidate genes fit with emerging data on atherosclerosis mechanisms. For example, a
218 knockout mouse for *ARHGEF26* on a hyperlipidemic background resulted in reduced
219 atherosclerosis and plaques with reduced macrophage content¹⁶. Similarly, *FN1* expression is
220 increased in plaques and mouse models have demonstrated a causal role for fibronectin-1 in
221 the development and progression of atherosclerosis¹⁷⁻¹⁸. Finally, we undertook a phenome
222 scan to assess pleiotropy (**Supplementary Table 10**). Several of the new lead SNPs (or a proxy)
223 had robust associations ($p < 5 \times 10^{-8}$) with traditional CAD risk factors such as LDL-cholesterol
224 (*HNF1A* and *FN1*), blood pressure (*PRDM8/FGF5*) and BMI (*SNRPD2*).

225

226 We next evaluated the broader functional relationships among genes associated with variants
227 ($N=11,427$) at 5%FDR. The 5%FDR set was annotated for eQTLs which, when present, were
228 mainly found in atherosclerotic aortic wall (25%) or internal mammary artery (22%) of CAD
229 patients (STARNET¹³; **Supplementary Table 9**). In GTEx¹⁵, eQTLs were mainly found in
230 subcutaneous fat (**Supplementary Table 9; Supplementary Figure 7**).

231 Prior pathway analyses of GWAS CAD loci have highlighted genes involved in lipid metabolism,
232 cellular movement, and processes such as tissue morphology and immune cell trafficking¹.

233 Analysis of 357 genes, selected as either eQTLs and/or the nearest gene to a 5%FDR
234 independent variant in this study (N=304), with the Ingenuity Knowledge base confirmed the
235 above findings¹ highlighting cardiovascular system development and function ($p = 1.31 \times 10^{-16}$),
236 organismal development ($p = 1.31 \times 10^{-16}$) and survival ($p = 1.52 \times 10^{-16}$) as the most
237 significant processes. In addition to canonical pathways related to lipid metabolism,
238 extracellular matrix, inflammation and nitric oxide production, the 357 gene set showed
239 enrichment for angiogenesis and signalling by the pro-angiogenic growth factor VEGF
240 (**Supplementary Figure 8**). We also applied DEPICT¹⁹ with the full distribution of 5%FDR
241 signals (**Online Methods**) to search for enriched gene sets. Blood vessel development, which
242 includes angiogenesis, was in the top 10 ($p < 6.67 \times 10^{-12}$) DEPICT Grouped-GeneSets
243 (GO:0001568; **Fig. 4, Supplementary Figure 9, Supplementary Table 11**).

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

247 This is the largest CAD genetic study to assess simultaneously common and rare (MAF <
248 1%)/low-frequency (MAF 1-5%) variants. In total, 101 low-frequency and 3 rare variants
249 reached genome-wide significance among all 5%FDR markers (N=11,427). This apparent
250 paucity in rare variants which has also been reported for type 2 diabetes²⁰, is likely due to lack
251 of power compared to studies of quantitative traits e.g. a study of adult height in ~700,000
252 individuals has reported 32 rare variants⁵. As expected, lower-frequency variants tend to have
253 stronger effects compared to common variants (**Supplementary Figure 10**) with the exception
254 of rs2891168 in *CDK2NB-AS1* (MAF 48.7%; OR 1.19; **Supplementary Table 13**). The intergenic
255 variant rs186696265 which had the largest OR (1.62) in our study is known to affect LDL
256 cholesterol levels²¹.

257 Our findings highlight the importance of the FDR approach to define an extended list of
258 associated variants. As we have previously proposed¹⁻², suggestive association signals in well-
259 powered GWAS such as this one can substantially improve our knowledge of disease
260 architecture at only a modest penalty implied by the 5%FDR. We have demonstrated the
261 potential value of the new 5%FDR list in improving prediction of CAD risk and implicating new
262 networks underlying CAD pathophysiology. This extended list of candidate genes provides a
263 powerful resource for functional studies.

264

265 **URLs**

266 www.ukbiobank.ac.uk/

267 [GWAS catalogue: https://www.ebi.ac.uk/gwas/](https://www.ebi.ac.uk/gwas/)

268 [GTEx portal: http://www.gtexportal.org/home/](http://www.gtexportal.org/home/)

269 [PhenoScanner: http://www.phenoscanter.medschl.cam.ac.uk/](http://www.phenoscanter.medschl.cam.ac.uk/)

270 [Ingenuity Knowledge Base: http://www.ingenuity.com/science/knowledge-](http://www.ingenuity.com/science/knowledge-)

271 [base?utm_source=Blog&utm_medium=link&utm_campaign=Doug%20Bassett%20ASHG%20](http://www.ingenuity.com/science/knowledge-base?utm_source=Blog&utm_medium=link&utm_campaign=Doug%20Bassett%20ASHG%20)

272 [2014](#)

273

274 **ACKNOWLEDGEMENTS**

275 This work was funded by British Heart Foundation (BHF) grants RG/14/5/30893 to P.D. and
276 FS/14/66/31293 to O.G. P.D. work forms part of the research themes contributing to the
277 translational research portfolio of Barts Cardiovascular Biomedical Research Unit which is
278 funded by the National Institute for Health Research (NIHR). F.Y.L. and S.E.H. are funded by
279 NIHR. C.P.N., T.R.W. and N.J.S. are funded by the BHF. N.J.S. is a NIHR Senior Investigator.
280 PROCARDIS was supported by EU-FP6 (LSHM-CT- 2007-037273), AstraZeneca, BHF, Swedish
281 Research Council, Knut and Alice Wallenberg Foundation, Swedish Heart-Lung Foundation,
282 Torsten and Ragnar Söderberg Foundation, Karolinska Institutet, Foundation Strategic
283 Research and Stockholm County Council (560283). M.F. and H.W. are supported by Wellcome
284 Trust award 090532/Z/09/Z; M.F., H.W. and T.K., the BHF Centre of Research Excellence. A.G,
285 H.W and T.K FP7/2007-2013 (HEALTH-F2-2013-601456 (CVGenes@Target),A.G. the
286 Wellcome Trust and TriPartite Immunometabolism Consortium- Novo Nordisk Foundation's
287 (NNF15CC0018486). HPS (ISRCTN48489393) was supported by Medical Research Council
288 (MRC), BHF, Merck and Co and Roche Vitamins Ltd. Merck and Co supported genotyping; HPS
289 acknowledges WT (07611) and National Blood Service and UK-Twins for controls. JCH is
290 supported by BHF (FS/14/55/30806). Mount Sinai BioMe Biobank is supported by The Andrea
291 and Charles Bronfman Philanthropies. GLACIER Study and P.W.F. is funded by Swedish
292 Research Council, Swedish Heart-Lung Foundation, Novo Nordisk, Umeå Medical Research
293 Foundation, and Swedish Diabetes Association. OHGS studies were funded by Canadian
294 Institutes of Health Research, Canada Foundation for Innovation and Heart & Stroke
295 Foundation of Canada. LURIC was funded from the EU-FP7 (Atheroremo (201668), RiskyCAD
296 (305739), INTERREG IV Oberrhein Program), European Regional Development Fund (ERDF),
297 Wissenschaftsoffensive TMO, and from the German ministry for education and research,
298 project e:AtheroSysMed (01ZX1313A-K). LOLIPOP is supported by NIHR-BRC Imperial College
299 Healthcare NHS Trust, BHF (SP/04/002), MRC (G0601966, G0700931), WT (084723/Z/08/Z),
300 NIHR (RP-PG-0407-10371), EU-FP7 (EpiMigrant, 279143) and Action on Hearing Loss (G51).
301 The Helsinki Sudden Death Study was funded by EU-FP7 (201668, AtheroRemo), Tampere
302 University Foundation, Tampere University Hospital Medical Funds (grants 9M048, 9N035 for
303 T.L), the Emil Aaltonen Foundation (T.L), Finnish Foundation of Cardiovascular Research (T.L,
304 P.K), Pirkanmaa Regional Fund of the Finnish Cultural Foundation, Yrjö Jahnsson Foundation,
305 Tampere Tuberculosis Foundation (T.L), Signe and Ane Gyllenberg Foundation (T.L.), and
306 Diabetes Research Foundation of Finnish Diabetes Association (T.L.).
307 The MRC/BHF Cardiovascular Epidemiology Unit was funded by MRC (G0800270), BHF
308 (SP/09/002), NIHR-BRC Cambridge, European Research Council (ERC; 268834), and EU-FP7
309 (HEALTH-F2-2012-279233), Pfizer, Merck and Biogen. EPIC-CVD was supported University of
310 Cambridge, EU-FP7 (HEALTH-F2-2012-279233), MRC (G0800270) BHF (SP/09/002), and ERC
311 (268834). We thank all EPIC participants and staff, Sarah Spackman for data management,
312 and EPIC-CVD Coordinating Centre team.(IUT20-60) and Centre of Excellence in Genomics and
313 Translational Medicine (GENTRANSMED), EU structural fund (Archimedes Foundation;
314 3.2.1001.11-0033), PerMed I and EU2020 (692145 ePerMed). This research was supported by
315 BHF (grant SP/13/2/30111) and conducted using the UK Biobank Resource (application
316 number 9922).

317
318 **AUTHOR CONTRIBUTIONS**

319 Writing group (wrote and edited manuscript): C.P.N., A.G., A.S.B., S.K., T.R.W., E.M., I.N.,
320 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,
321 and commented on the manuscript. Data generation & cohorts: A.S.B., O.G., T.J., L.Z., S.E.H.,
322 E.A., T.L.A., E.P.B., J.C.C., R.C., P.C., R.M.C., R.E., E.E., P.W.F., C.G., D.G., A.H., J.M.M.H., E.I.,
323 A.K., T.K., T.K., T.L., X.L., Y.L., W.M., R.McP., A.M., M.Pujades-R., A.F.S., M.J.S., P.A.Z., R.J.F.L.,
324 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.,
325 A.S.B., I.N., 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.,
326 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:
327 C.P.N., S.K., T.W., A.S.B., R.E., A.R., E.E.S., J.L.M.B. Biological and clinical enrichment and
328 pathway analyses: E.M., P.D.

329

330

331 **COMPETING FINANCIAL INTERESTS**

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

342

343

344 **References**

- 345 1. CARDIoGRAMplusC4D Consortium, Deloukas, P., Kanoni, S., Willenborg, C. *et al.* Large-scale
346 association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45**, 25-33
347 (2013).
- 348 2. CARDIoGRAMplusC4D Consortium. A comprehensive 1,000 Genomes-based genome-wide
349 association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121-1130 (2015).
- 350 3. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Coding
351 Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N. Engl. J. Med.* **374**,
352 1134-1144 (2016)
- 353 4. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators.
354 Systematic evaluation of pleiotropy identifies six further loci associated with coronary artery
355 disease. *J. Am. Col. Cardiol.* **69**, 823-836 (2017)
- 356 5. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature*
357 **542**, 186-190 (2017)
- 358 6. Cardiovascular Disease Statistics 2015. British Heart Foundation
- 359 7. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E. *et al.* Conditional and joint multiple-SNP
360 analysis of GWAS summary statistics identifies additional variants influencing complex traits.
361 *Nat. Genet.* **44**, 369-375, S361-363 (2012)
- 362 8. Danesh, J., Saracci, R., Berglund, G., Feskens, E. *et al.* EPIC-Heart: the cardiovascular
363 component of a prospective study of nutritional, lifestyle and biological factors in 520,000
364 middle-aged participants from 10 European countries. *Eur. J. Epidemiol.* **22**, 129-141 (2007)
- 365 9. Wipff, P.J., Rifkin, D.B., Meister, J.J., Hinz, B. Myofibroblast contraction activates latent TGF-
366 beta1 from the extracellular matrix. *J. Cell Biol.* **179**, 1311-23(2007)
- 367 10. Henderson, N.C., Arnold, T.D., Katamura, Y., Giacomini, M.M. *et al.* Targeting of α v integrin
368 identifies a core molecular pathway that regulates fibrosis in several organs. *Nat. Med.* **19**,
369 1617-24 (2013)
- 370 11. Goumans, M.J., Liu, Z., ten Dijke, P. TGF-beta signaling in vascular biology and dysfunction.
371 *Cell Res.* **19**, 116-27 (2009)
- 372 12. Nanda, V., Downing, K.P., Ye, J., Xiao, S. *et al.* CDKN2B Regulates TGF β Signaling and
373 Smooth Muscle Cell Investment of Hypoxic Neovessels. *Circ. Res.* **118**, 230-40 (2016).
- 374 13. Franzén, O., Ermel, R., Cohain, A., *et al.* Cardiometabolic risk loci share downstream cis-
375 and trans-gene regulation across tissues and diseases. *Science* **353**, 827-830 (2016)

- 376 14. Kukreja, R.C., Salloum, F.N., Das, A. Cyclic guanosine monophosphate signaling and
377 phosphodiesterase-5 inhibitors in cardioprotection. *J. Am. Coll. Cardiol.* **59**, 1921–
378 1927 (2012)
- 379 15. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multi-tissue
380 gene regulation in humans. *Science* **348**, 648-660
- 381 16. Samson, T., van Buul, J.D., Kroon, J., Welch C, *et al.* The guanine-nucleotide exchange
382 factor SGEF plays a crucial role in the formation of atherosclerosis. *PLoS One* **8**, e55202 (2013)
- 383 17. Babaev, V.R., Porro, F., Linton, M.F., Fazio, S. *et al.* Absence of regulated splicing of
384 fibronectin EDA exon reduces atherosclerosis in mice. *Atherosclerosis* **197**, 534-540 (2008)
- 385 18. Rohwedder, I., Montanez, E., Beckmann, K., Bengtsson, E. *et al.* Plasma fibronectin
386 deficiency impedes atherosclerosis progression and fibrous cap formation. *EMBO Mol. Med.*
387 **4**, 564-576 (2012)
- 388 19. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using
389 predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
- 390 20. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-47 (2016)
- 391 21. Surakka, I. *et al.* The impact of low-frequency and rare variants on lipid levels. *Nat.*
392 *Genet.* **47**, 589-97 (2015)
- 393

394 **Figure legends**

395 **Figure 1. (a)** Diagram depicting the CAD phenotype definition in UK Biobank. HARD CAD
396 defined as fatal or non-fatal myocardial infarction (MI), PTCA (percutaneous transluminal
397 coronary angioplasty), or coronary artery bypass grafting (CABG). SOFT CAD includes HARD
398 CAD as well as chronic ischaemic heart disease (IHD) and angina. UK Biobank self-reported
399 data: 'Vascular/heart problems diagnosed by doctor' or 'Non-cancer illnesses that self-
400 reported as angina or heart attack'. Self-reported surgery defined as either PTCA, CABG or
401 triple heart bypass. HESIN hospital episodes data and death registry data using diagnosis and
402 operation - primary and secondary cause: MI defined as hospital admission or cause of death
403 due to ICD9 410-412, ICD10 I21-I24, I25.2; PTCA is defined as hospital admission for PTCA
404 (OPCS-4 K49, K50.1, K75); CABG is defined as hospital admission for CABG (OPCS-4 K40 –
405 K46); Angina or chronic IHD defined as hospital admission or death due to ICD9 413, 414.0,
406 414.8, 414.9, ICD10 I20, I25.1, I25.5-I25.9. **(b)** Radar plot highlighting the proportions (%) of
407 signals between the HARD and SOFT CAD phenotype definitions based on the 5%FDR results
408 **(Supplementary Table 4)**; MAF = minor allele frequency, $p < 5 \times 10^{-8}$ marks variants reaching
409 genome-wide significance, OR = odds ratio (OR > 1.05 corresponds to 85% power to detect a
410 signal ($\alpha < 0.05$) in the SOFT analysis). The results for all six subgroups of variants assessed
411 did not differ statistically between the two phenotype definitions ($p > 0.1$)

412 **Figure 2.** Transposed Manhattan plot showing the SOFT meta-analysis results under an
413 additive model. The P -values are truncated at $-\log_{10}(P) = 20$. The red dotted lines are at
414 GWAS ($P = 5 \times 10^{-8}$) and 5% FDR significance ($P = 6.28 \times 10^{-5}$). The known CAD risk loci are shown
415 in black **(Supplementary Table 2)**; *KSR2* and *ZNF507-LOC400684* had reached genome-wide
416 significance under a recessive model². The exome chip markers are shown with an *. The 13

417 novel CAD loci which reached genome-wide significance in our study (including replication
418 data; **Table 1**), are written in brown font.

419 **Figure 3.** Single marker p-value comparison of the 5% FDR variants in the published
420 CARDIoGRAMplusC4D 1000Genomes CAD GWAS meta-analysis² and current FDR study. Of
421 the 162 variants which had $p < 5 \times 10^{-5}$ in the CAD 1000Genomes GWAS, 116 had a match or
422 good proxy ($r^2 > 0.8$) in the new FDR list (red circles). SNPs in green (n=7) were present in the
423 earlier FDR list and reached genome-wide significance in the current analysis.

424 **Figure 4.** Heat map showing the DEPICT gene set enrichment results with zoom-in on a subset
425 of the results. 556 gene sets are included which had evidence of enrichment at 1% FDR. The
426 x– axis shows the gene name, which is predicted to be included in the reconstituted gene set
427 indicated in the y – axis. The color red indicates higher Z-score, where Z-score is a value
428 representing each gene’s inclusion in DEPICT’s reconstituted gene sets. Clustering was made
429 based on complete linkage method. Highlighted pathways in the cluster, include
430 angiogenesis, blood vessel development and morphogenesis.

431

432 **Online Methods**

433 **Phenotype Definitions & Power calculation**

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

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

444 Susceptibility effect sizes in MI cases and an inclusive CAD definition were very similar in
445 the earlier GWAS². We hypothesized that the detailed clinical information in UKBB might
446 enhance the search for novel loci by further broadening the CAD phenotype to increase
447 sample size.

448

449 **GWAS and meta-analyses**

450 All participants gave written consent for participation in genetic studies, and the protocol
451 of each study was approved by the corresponding local research ethics committee or
452 institutional review board. Participating cohorts in the 1000 Genomes and Exome GWAS
453 studies are described elsewhere^{2,3}. UK Biobank (UKBB samples) were excluded due to
454 withdrawn consent, sex mismatches (n=182), Biobank/Believe QC exclusions (n=406) and
455 sample relatedness (n=3,481) determined as $\text{Kinship} > 0.088$. GWAS analysis in UKBB was
456 restricted to variants with results available in the published GWAS² or Exome³⁻⁴ dataset.
457 Further exclusions included poorly imputed ($\text{info} < 0.4$) or monomorphic variants, duplicate
458 variants across data sets, variants that deviated strongly from Hardy-Weinberg Equilibrium
459 in European ancestry controls ($p < 1 \times 10^{-9}$), variants with an effect allele frequency in
460 European ancestry samples that differed strongly (i) from 1000G European panel, (ii) from
461 GWAS/Exome data, (iii) between arrays (UKBB vs UK-BiLEVE), and (iv) across genotyping
462 batches. Variants that did not produce a valid result or estimated extreme log odds ratios
463 ($|\beta| > 4$) were also excluded after analysis. Cluster plots lead variants and of proxies
464 were visually inspected.

465 We ran the GWAS under an additive frequentist mode of inheritance for each variant using
466 the dosages from the imputed data, adjusting for array (UK Biobank vs UK BiLEVE) and the
467 first five principal components using SNPTEST. Age and sex were not adjusted for to
468 maximize the power to detect associations with diseases that have a prevalence $< 10\%$ ²².
469 Population stratification was assessed and standard errors were adjusted using the
470 genomic inflation statistic (λ).

471 Association summary statistics (after λ correction) from the UKBB were combined with the
472 1000 Genomes (1000G) imputed GWAS results² and the Exome results³ via two separate
473 fixed-effect inverse-variance weighted meta-analysis implemented in GWAMA²³. We
474 applied post meta-analysis λ correction in each instance. We identified 36,460 variants
475 present in both the 1000G imputed GWAS and the Exome results. We retained the variants
476 from the 1000G imputed GWAS if the median info score was 1, otherwise we retained the
477 results from the Exome data.

478

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

480 The false discovery rate (FDR) following the meta-analysis with UKBB was assessed using a
481 step-up procedure in the *qqvalue* Stata program²⁴ as it is well controlled under positive
482 regression-dependency conditions. We used the Simes method to generate q-values for
483 the 8.9M variants. The p-value cut-off for a q-value of 5% for HARD was 7.24×10^{-5} and SOFT
484 was 6.28×10^{-5} . Peak SNPs were identified in a 1cM window. There is an exact overlap of
485 155 variants between the 2 peak variant lists, however, using the 1cM window the overlap
486 increases to 206 variants. Both the lists were annotated and classified into 6 categories
487 (exome chip, indels, Odds Ratio (OR) > 1.05 , $p < 5 \times 10^{-8}$, MAF $< 5\%$ and exonic). The proportions
488 were calculated in each of the 6 categories and plotted as a radar plot (**Fig. 1b**). Monte
489 Carlo simulations were used to assess the *post-hoc* power of the UKBB interim data to
490 replicate the 155 variants. The 1000G GWAS effect sizes (“betas”) are expected to be
491 subject to *winner’s curse* inflation so were shrunken (towards the null) by application of
492 the FIQT procedure²⁵. Effect sizes for firmly established CAD loci were systematically lower
493 for SOFT compared to the HARD phenotype (**Supplementary Table 1**) noting that HARD
494 closely corresponds to the CAD phenotype in reference 2. Betas were therefore further

495 shrunken by a factor $\log(1.059)/\log(1.072) = 0.82$ (**Supplementary Table 1**). 10,000
496 replicates were then randomly drawn from the vector of shrunken betas and the
497 corresponding UKBB standard errors, to allow for variation in genotype call rates,
498 imputation quality and allele frequency and to calculate Wald association statistics.
499 Multiple testing of 155 variants was allowed for by controlling the FDR to 5% with a step-
500 up procedure encoded in the *multproc*²⁶ Stata™ program. The average expected number
501 of replicated variants was 56 (95%CI 42 – 69). Testing the 5% FDR variants (**Supplementary**
502 **Table 7**) in UKBB with a model adjusted for age and sex gave concordant results to the
503 unadjusted model (data not shown).

504

505 **GCTA & Heritability analysis**

506 We used the GCTA software⁷ to perform joint association analysis in (SOFT) meta-analysis
507 results. This approach fits an approximate multiple regression model using summary-level
508 meta-analysis statistics and LD corrections estimated from a reference panel (here the
509 UKBB sample). We adopted a chromosome-wide stepwise selection procedure to select
510 variants and estimate their joint effects at i) a genome-wide significance level ($p_{\text{Joint}} \leq$
511 5×10^{-8}) in the totality of meta-analysed variants ($n \sim 9\text{M}$; **Supplementary Figure 10**,
512 **Supplementary Table 11**) and ii) a Bonferroni-corrected $p_{\text{Joint}} < 1 \times 10^{-4}$ corresponding to
513 the number of independent LD bins ($r^2 < 0.1$) in the 5% FDR variant list ($n=11,427$;
514 **Supplementary Table 6**).

515 Heritability calculations were based on a multifactorial liability-threshold model,
516 implemented in the INDI-V²⁷ calculator (<http://cnsgenomics.com/shiny/INDI-V/>), under
517 the assumption of a baseline population risk (K) of 0.0719²⁸ and a twins heritability (H_L^2) of
518 0.4. Multiple regression estimates from the GCTA joint association analysis were used to
519 estimate heritability for the 304 independent CAD risk variants within the 5% FDR list.

520

521 **Genetic risk score analysis**

522 GRS analysis was undertaken in the EPIC-CVD study⁸ which comprises 7910 CAD cases and
523 12958 controls (**Supplementary Note**). We considered either all known and new lead CAD
524 risk variants reaching genome-wide significance (GRS2; **Supplementary Table 2** and **Table**
525 **1**) or the 304 variants in the 5% FDR set (GRS1; **Supplementary Table 7**). We used variants
526 with an INFO score filter of 0.4 in EPIC-CVD and replaced missing ones with proxies ($r^2 >$
527 0.8 in 1000 Genomes European participants). GRS1 comprised 280 variants and GRS2 71.
528 The raw GRS was obtained by summing the dosages of these variants for all individuals.
529 We then fitted a Prentice weighted cox regression model for each GRS, adjusting for age
530 and sex, to obtain survival forecasts and calculate the C indices. Statistical analyses were
531 performed using R 3.3.3 and STATA 13.1. Variant extraction was done using qctool 1.4.

532

533 **Functional annotation**

534 **eQTLs:** For associations between the 304 independent variants (5% FDR) and gene
535 expression traits we searched for expression quantitative trait loci (eQTLs) in the
536 Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET) RNA-seq
537 dataset¹³ and the Genotype-Tissue Expression¹⁵ (GTEx) portal. eQTLs were included if the
538 best eSNP (i.e. the variant with the most significant association with gene expression in cis)
539 was in high LD ($r^2 > 0.8$) with the CAD lead SNP.

540 **Regulatory elements:** We functionally annotated each of the 13 lead variants and their
541 proxies ($r^2 > 0.8$) using HaploregV4²⁹. Overlap with regulatory elements including
542 chromosome state segmentation, DNase hypersensitivity, and transcription factor binding
543 (TFB) as determined by the ENCODE³⁰ and Roadmap Epigenome projects³¹, and predicted
544 effects on TFB based on regulatory motifs from TRANSFAC³² and JASPAR³³ were identified
545 using HaploregV4¹⁹ and the UCSC genome browser. Variants were then scored using three
546 different bioinformatics tools that help prioritise causal disease variants. Combined
547 Annotation Dependent Depletion (CADD)³⁴ incorporates a range pathogenicity prediction
548 tools to provide a genome-wide score (C-score) for each test variant from its pre-calculated
549 database of ~8.6 billion genetic variants. High scores indicate variants that are not
550 stabilized by selection and are more likely to be disease-causing and low scores indicate
551 evolutionary stable non-damaging variants. The top 10% of likely functional variants will
552 have a C-score > 10 and top 1% of variants will have a C-score > 20 . Genome-wide
553 annotation of variants (GWAVA)³⁵ predicts the functional impact of noncoding variants
554 based on genomic and epigenomic annotations and provides scores between 0 and 1 with
555 higher scores indicating variants that are more likely to be functional. RegulomeDB³⁶
556 annotates and scores variants in seven categories based datasets such as ENCODE. Scores
557 of 1-2 variants likely to affect TFB, 3 less likely to affect binding, 4-6 relate to variants with
558 minimal binding evidence and 7 is for variants with no regulatory annotation.

559 **Phenome-scan:** look ups in other common traits were performed using the PhenoScanner
560 database as described in ref 37.

561

562 **Pathway analysis**

563 **DEPICT:** DEPICT¹⁹ is a computational tool which performs gene set enrichment analyses to
564 prioritize genes in associated GWAS loci with probabilistically predefined gene sets based
565 on Gene Ontology terms, canonical pathways, protein-protein interaction subnetworks
566 and rodent phenotypes; reconstituted gene sets are detailed in refs 19 and 38. Input to
567 our analysis were the 11,427 CAD variants (FDR 5%) of which 11,311 were annotated in
568 DEPICT. We constructed loci as previously described (beta version 1.1, release 194,
569 www.broadinstitute.org/mpg/depict). Analysis was performed with default parameters
570 (50 repetitions to compute FDRs, 500 permutations to adjust for biases, such as gene
571 length). The 11,311 variants were collapsed to 288 loci which were used in the gene set
572 enrichment analyses. Correlated gene sets were grouped together based on gene
573 membership to expedite data interpretation.

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

586 of query molecules that participate in that function and the total number of molecules that
587 are known to be associated with that function in the IKB.

588

589

590 **Data Availability:** Meta-analysis summary statistics for the variants considered in this
591 study for association with CAD (SOFT definition) are available at
592 <http://www.cardiogramplusc4d.org/data-downloads/>.

593

594

595

596

597 **Method References**

- 598 22. Pirinen, M., Donnelly, P. & Spencer, CC. Including known covariates can reduce power
599 to detect genetic effects in case-control studies. *Nat Genet* **44**, 848-51 (2012).
- 600 23. Magi, R. & Morris, AP. GWAMA: software for genome-wide association meta-analysis.
601 *BMC Bioinformatics*, **11**:288 (2010).
- 602 24. Newson, R. QQVALUE: Stata module to generate quasi-q-values by inverting multiple-
603 test procedures, Statistical Software Components S457100, Boston College Department
604 of Economics, (2009, revised 2013)
- 605 25. Bigdeli, T.B. *et al.* A simple yet accurate correction for winner's curse can predict
606 signals discovered in much larger genome scans. *Bioinformatics* **32**, 2598-603 (2016)
- 607 26. Newson R. Multiple-test procedures and smile plots. *The Stata Journal* **3**, 109-132
608 (2003).
- 609 27. Witte, J.S., Visscher, P.M. & Wray, N.R. The contribution of genetic variants to disease
610 depends on the ruler. *Nat Rev Genet* **15**, 765-76 (2014).
- 611 28. Muñoz, M. *et al.* Evaluating the contribution of genetics and familial shared
612 environment to common disease using the UK Biobank. *Nat Genet* **48**, 980-3 (2016).
- 613 29. Ward, L.D., Kellis, M. HaploReg: a resource for exploring chromatin states,
614 conservation, and regulatory motif alterations within sets of genetically linked variants.
615 *Nucleic Acids Res.* **40**, D930-4 (2012).
- 616 30. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the
617 human genome. *Nature* **489**, 57-74 (2012).
- 618 31. Skipper, M., Eccleston, A., Gray, N., *et al.* Presenting the epigenome roadmap. *Nature*
619 **518**, 313 (2015).
- 620 32. Matys, V., Fricke, E., Geffers, R., *et al.* TRANSFAC: transcriptional regulation, from
621 patterns to profiles. *Nucleic Acids Res* **31**:374–8 (2003).
- 622 33. Portales-Casamar, E., Thongjuea, S., Kwon, A.T., *et al.* JASPAR 2010: the greatly
623 expanded open-access database of transcription factor binding profiles. *Nucleic Acids*
624 *Res.* **38**, D105–10 (2010).
- 625 34. Kircher, M., Witten, D.M., Jain, P., *et al.* A general framework for estimating the
626 relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
- 627 35. Ritchie, G.R., Dunham, I., Zeggini, E., *et al.* Functional annotation of noncoding
628 sequence variants. *Nat Methods* **11**, 294-6 (2014).
- 629 36. Boyle, A.P., Hong, E.L., Hariharan, M., *et al.* Annotation of functional variation in
630 personal genomes using RegulomeDB. *Genome Research* **22**, 1790-7 (2012)
- 631 37. Staley, J.R. *et al.* PhenoScanner: a database of human genotype–phenotype
632 associations. *Bioinformatics* **32**, 3207–3209 (2016)

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

635

636

637

638 **Table 1** - Novel variants reaching genome-wide significance ($P < 5 \times 10^{-8}$) in [the combined \(discovery and replication\) SOFT meta-analysis](#)

Locus Name	Markername	CHR	POS (hg19)	EA	EAF	Functional Evidence	UKBB+CoG/Exome			Meta analysis	
							OR (95% CI)	P-value	FDR Q-value	OR (95% CI)	P-value
TDRKH	rs11810571	1	151762308	G	0.849	eQTL/coding	1.060 (1.039, 1.082)	2.21×10^{-8}	8.05×10^{-5}	1.057 (1.036, 1.079)	4.24×10^{-8}
FN1	rs1250229*	2	216304384	T	0.256	eQTL/coding	1.072 (1.052, 1.092)	1.85×10^{-13}	2.05×10^{-9}	1.071 (1.051, 1.091)	2.77×10^{-13}
RHOA	rs7623687	3	49448566	A	0.855	none	1.074 (1.049, 1.100)	3.72×10^{-9}	1.62×10^{-5}	1.076 (1.052, 1.101)	3.44×10^{-10}
UMPS/ITGB5	rs142695226	3	124475201	G	0.138	eQTL/coding	1.069 (1.045, 1.094)	1.00×10^{-8}	3.98×10^{-5}	1.071 (1.048, 1.095)	1.53×10^{-9}
ARHGEF26	rs12493885*	3	153839866	C	0.886	eQTL	1.074 (1.047, 1.101)	3.29×10^{-8}	1.15×10^{-4}	1.073 (1.047, 1.101)	3.16×10^{-8}
PRDM8/FGF5	rs10857147	4	81181072	T	0.275	none	1.056 (1.036, 1.075)	8.96×10^{-9}	3.60×10^{-5}	1.054 (1.036, 1.073)	5.66×10^{-9}
PDE5A/MAD2L1	rs7678555	4	120909501	C	0.30	eQTL	1.049 (1.031, 1.069)	1.43×10^{-7}	4.25×10^{-4}	1.052 (1.034, 1.070)	1.32×10^{-8}
HDGFL1	rs6909752	6	22612629	A	0.351	none	1.051 (1.034, 1.069)	5.59×10^{-9}	2.35×10^{-5}	1.051 (1.034, 1.068)	2.19×10^{-9}
ARNTL	rs3993105	11	13303071	T	0.704	none	1.048 (1.030, 1.067)	1.06×10^{-7}	3.33×10^{-4}	1.048 (1.031, 1.066)	4.77×10^{-8}
HNF1A	rs2244608	12	121416988	G	0.355	coding	1.053 (1.035, 1.070)	2.32×10^{-9}	1.06×10^{-5}	1.053 (1.035, 1.070)	7.74×10^{-10}
CDH13	rs7500448	16	83045790	A	0.752	eQTL	1.061 (1.040, 1.082)	5.14×10^{-9}	2.18×10^{-5}	1.063 (1.043, 1.083)	4.76×10^{-10}
TGFB1	rs8108632	19	41854534	T	0.488	none	1.049 (1.031, 1.067)	5.88×10^{-8}	1.95×10^{-4}	1.048 (1.031, 1.066)	4.04×10^{-8}
SNRPD2	rs1964272	19	46190268	G	0.510	none	1.045 (1.028, 1.063)	2.29×10^{-7}	6.15×10^{-4}	1.047 (1.030, 1.064)	2.46×10^{-8}

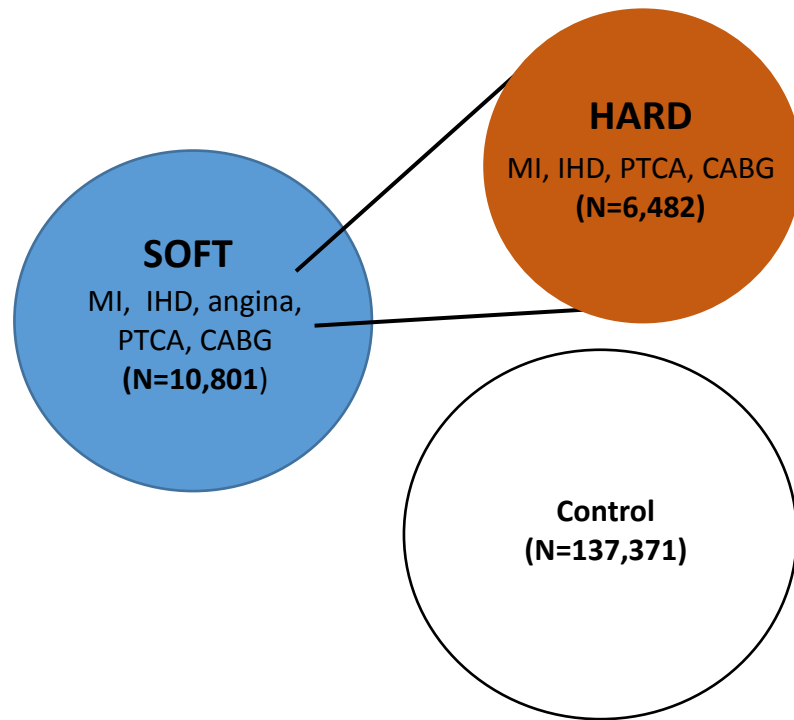
639 *Exome marker

640 [EA: effect allele; EAF: Effect allele frequency; CoG = CARDIoGRAMplusC4D 1000G GWAS; Exome = Exome array analysis; UKBB = UK Biobank;](#)
641 [Discovery sample comprised 71,602 cases and 260,875 controls \(for exome markers 53,135 and 215,611 respectively\); Replication sample](#)
642 [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 ($r^2 > 0.8$)
643 is a coding change, has evidence as an expression quantitative trait locus (eQTL), or both. Further details of functional evidence are provided in
644 **Supplementary Table 7** and **Supplementary Figure 6**.
645

646

Figure 1

a



b

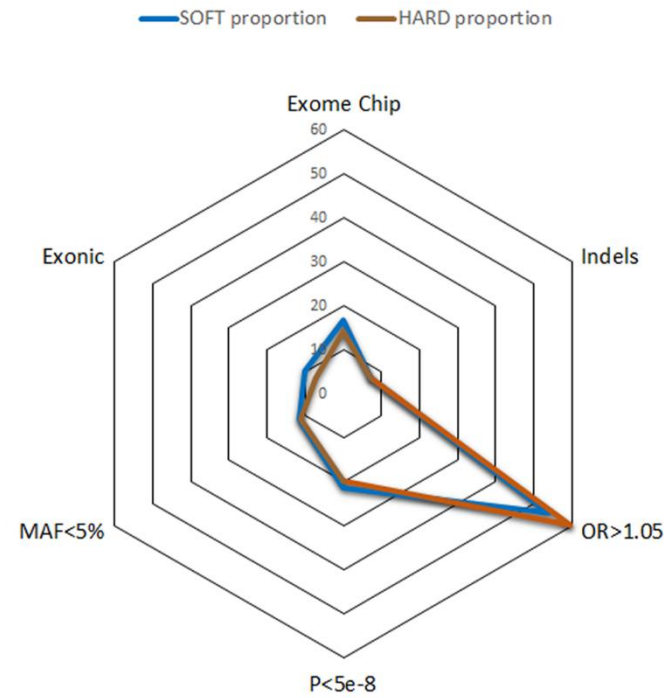


Figure 2

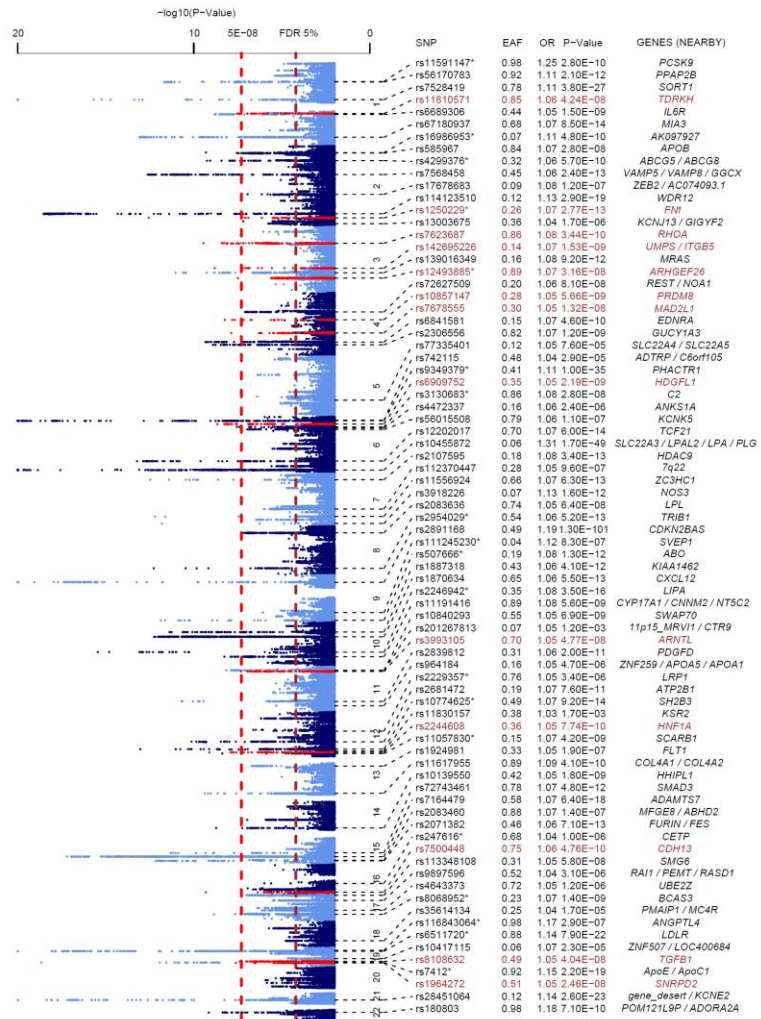


Figure 3

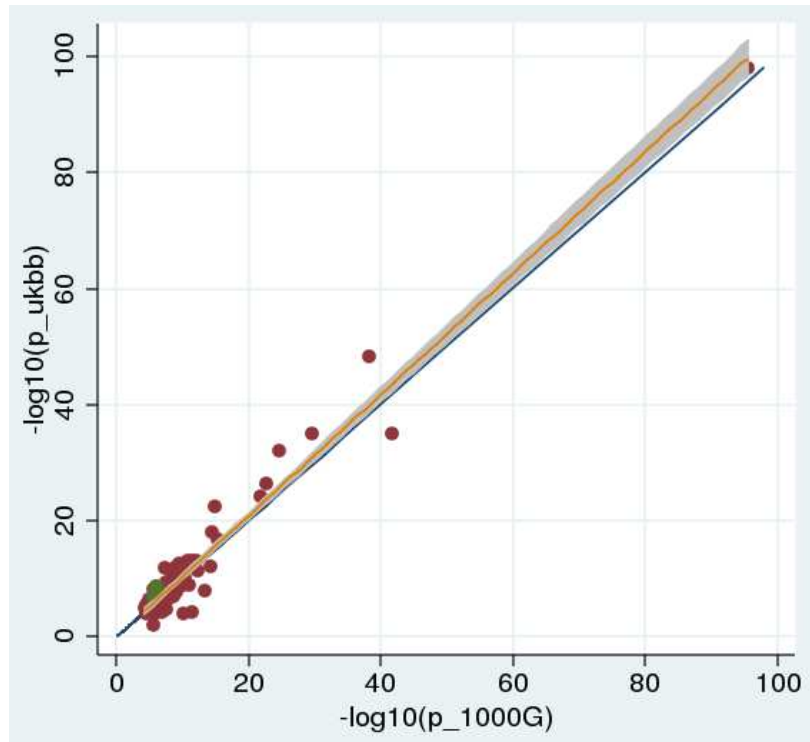


Figure 4

