Genetic loci associated with prevalent and incident myocardial infarction and coronary heart disease in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium

Authors

Julie Hahn[*1], Yi-Ping Fu[*2,3], Michael R. Brown*[1], Joshua C. Bis[*4], Paul S. de Vries[*1,5], Mary F. Feitosa[*6], Lisa R. Yanek[*7], Stefan Weiss[*8,9], Franco Giulianini[*10], Albert Vernon Smith[*11,12], Xiuqing Guo[*13], Traci M. Bartz[4,14], Diane M. Becker[7], Lewis C. Becker[7], Eric Boerwinkle[1,15], Jennifer A. Brody[4], Yii-Der Ida Chen[13], Oscar H. Franco[16], Megan Grove[1], Tamara B. Harris[17], Albert Hofman[16,18], Shih-Jen Hwang[3], Brian G. Kral[7], Lenore J. Launer[16], Marcello R. P. Markus[9, 19], Kenneth M. Rice[14], Stephen S. Rich[20], Paul M. Ridker[10,21], Fernando Rivadeneira[16], Jerome I. Rotter[13], Nona Sotoodehnia[22], Kent D. Taylor[13], André G. Uitterlinden[16,23], Uwe Völker[8,9], Henry Völzke[9,24], Jie Yao[13], Daniel I. Chasman[10,21], Marcus Dörr[9,19], Vilmundur Gudnason[11,12], Rasika A. Mathias[7], Wendy Post[20], Bruce M. Psaty[4,25,26,27], Abbas Dehghan[5,18], Christopher J. O'Donnell[**3,21,28,29], Alanna C. Morrison[**1]

* authors equally shared first author contributions and responsibilities

** authors equally shared senior author contributions and responsibilities

Affiliations

[1] Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas USA

[2] Office of Biostatistics Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA

[3] Framingham Heart Study, National Heart, Lung, and Blood Institute, National Institutes of Health, Framingham, Massachusetts, USA

27 [4] Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle,
28 Washington, USA
29 [5] Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands
30 [6] Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine,
31 St. Louis, Missouri, USA
32 [7] GeneSTAR Research Program, Department of Medicine, Johns Hopkins University School of Medicine,
33 Baltimore, Maryland, USA
34 [8] Interfaculty Institute for Genetics and Functional Genomics, The University Medicine and Ernst-Moritz-
35 Arndt-University Greifswald, Greifswald, Germany
36 [9] DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, Germany
37 [10] Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA
38 [11] Icelandic Heart Association, Kovapvogur, Iceland
39 [12] Faculty of Medicine, University of Iceland, Reykajvik, Iceland
40 [13] The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The
41 Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, California,
42 USA
43 [14] Department of Biostatistics, The University of Washington, Seattle, Washington, USA
44 [15] Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA
45 [16] Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands
46 [17] Laboratory of Epidemiology and Population Sciences, Intramural Research Program, National Institute
47 on Aging, National Institutes of Health, Bethesda, Maryland, USA
48 [18] Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom
49 [19] Department of Internal Medicine B - Cardiology, Pneumology, Infectious Diseases, Intensive Care
50 Medicine, The University Medicine Greifswald, Greifswald, Germany
51 [20] Department of Medicine and Epidemiology, Johns Hopkins University School of Medicine, Baltimore,
52 Maryland, USA

53 [21] Harvard Medical School, Boston, Massachusetts, USA

54 [22] Division of Cardiology, Department of Medicine, University of Washington, Seattle, Washington, USA

55 [23] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts,

56 USA

57 [24] Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

58 [25] Department of Epidemiology, The University of Washington, Seattle, Washington, USA

59 [26] Department of Health Services, The University of Washington, Seattle, Washington, USA

60 [27] Kaiser Permanente Research Institute, Seattle, Washington, USA

61 [28] VA Boston Healthcare System, Veteran's Affair, Boston, Massachusetts, USA

62 [29] Cardiovascular Medicine Division, Brigham and Women's Hospital, Boston, Massachusetts, USA

Abstract

Background

Genome-wide association studies have identified multiple genomic loci associated with coronary artery disease, but most are common variants in non-coding regions that provide limited information on causal genes and etiology of the disease. To better understand etiological pathways that might lead to discovery of new treatments or prevention strategies, we focused our investigation on low-frequency and rare sequence variations primarily residing in coding regions of the genome while also exploring associations with common variants.

Methods and Results

Using samples of individuals of European ancestry from ten cohorts within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, both cross-sectional and prospective analyses were conducted to examine associations between genetic variants and myocardial infarction (MI), coronary heart disease (CHD), and all-cause mortality following these events. Single variant and gene-based analyses were performed separately in each cohort and then meta-analyzed for each outcome. A low-frequency intronic variant (rs988583) in *PLCL1* was significantly associated with prevalent MI (OR=1.80, 95% confidence interval: 1.43, 2.27; P=7.12 × 10^{-7}). Three common variants, rs9349379 in *PHACTR1*, and rs1333048 and rs4977574 in the 9p21 region, were significantly associated with prevalent CHD. Four common variants (rs4977574, rs10757278, rs1333049, and rs1333048) within the 9p21 locus were significantly associated with incident MI. We conducted gene-based burden tests for genes with a cumulative minor allele count (cMAC) $\geq$ 5 and variants with minor allele frequency (MAF) < 5%. *TMPRSS5* and *LDLRAD1* were significantly associated with prevalent MI and CHD, respectively, and *RC3H2* and *ANGPTL4* were significantly associated with incident MI and CHD, respectively. No loci were significantly associated with all-cause mortality following a MI or CHD event.

Conclusion

87 This study confirmed previously reported loci influencing heart disease risk, and one single variant and

88 three genes associated with MI and CHD were newly identified and warrant future investigation.

Introduction

Coronary heart disease (CHD) is a leading cause of morbidity and mortality worldwide, accounting for one of every seven deaths in the United States in 2016. (1) In addition to major modifiable risk factors such as dyslipidemia, hypertension, diabetes, and cigarette smoking (2), genetic susceptibility to CHD has also been investigated extensively through family-based studies, candidate gene studies, and more recently genome-wide association studies (GWAS). (3-9) With progressively expanded sample sizes in recent GWAS, at least 160 loci have been associated with the risk of coronary artery disease. (10-13) Most of these loci are represented by common variants located in noncoding regions, resulting in limited implications for causal genes and etiological pathways. Further, while most available data are derived from genome-wide analysis of prevalent CHD, data are sparse from prospective studies of incident cardiovascular events in populations free of baseline cardiovascular disease.

Low-frequency and rare coding sequence variations across the genome have been investigated in studies of cardiovascular disease risk factors (14-18), with the goal of better understanding the etiology of these risk factors and to advance the discovery of the treatment and prevention of diseases. (19) We previously published the results from a prospective analysis of CHD among individuals of European ancestry from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, and identified low-frequency and common variants associated with incident CHD. (20)

In this current study of individuals of European ancestry, we implemented both a cross-sectional and prospective study design in the setting of the CHARGE Consortium to examine the association between genetic variants and the risk of prevalent and incident myocardial infarction (MI) and CHD. Study of incident cardiovascular events is enabled by the rigorous prospective design of population cohorts contributing to the CHARGE Consortium. We also investigated whether these genetic variants are associated with all-cause mortality after incident MI and CHD.

Materials and Methods

Study design and participants

Ten cohorts within the CHARGE Consortium Subclinical Working Group were included in this study: Age, Gene, Environment, Susceptibility Study (AGES), Atherosclerosis Risk in Communities (ARIC) Study, Cardiovascular Health Study (CHS), Family Heart Study (FamHS), Framingham Heart Study (FHS), the GeneSTAR Study (GeneSTAR), Multi-Ethnic Study of Atherosclerosis (MESA), Rotterdam Study (RS), Study of Health in Pomerania (SHIP), and the Women's Genome Health Study (WGHS). Detailed characteristics of the participating cohorts and study participant are shown in the **Supporting Document.** All study participants provided written informed consent to participate in genetic studies, and all study sites received approval to conduct this research from their local Institutional Review Boards (IRB) respectively.

Genotype calling and quality control

Participants from WGHS were genotyped by the HumanHap300 Duo+ (Illumina, Inc., San Diego, CA), and all other study participants were genotyped by the HumanExome BeadChip (v1.0-1.2, Illumina, Inc., San Diego, CA) which contains more than 240,000 variants including those discovered through exome sequencing in ~12,000 individuals and other non-coding common variants such as previously-reported GWAS signals and ancestry-informative markers. Data for AGES, ARIC, CHS, FamHS, FHS, MESA, and RS were jointly called at the University of Texas Health Science Center at Houston (21); SHIP was called in Illumina GenomeStudio using the CHARGE Consortium joint calling cluster file; GeneSTAR used the Illumina GenomeStudio and zCall software (22); and WGHS data was called using the Illumina BeadStudio v.3.3. Variant quality control (QC) was performed centrally (21) and by the individual studies, including checking concordance with previous GWAS data, and excluding participants with missing >5% genotypes, population clustering outliers, individuals with high inbreeding coefficients or heterozygote rates, gender mismatches, duplicated pairs, and unexpectedly high proportion of identity-by-descent sharing for family studies.

Cardiovascular outcome definition

Two cardiovascular outcomes were examined for association in this study: 1) MI: fatal or non-fatal MI; and 2) CHD: fatal or non-fatal MI, fatal CHD, sudden death within one hour of onset of symptoms, or revascularization (percutaneous coronary artery intervention such as stent or balloon angioplasty, or coronary artery bypass grafting). No exclusions were applied for the cross-sectional analysis of prevalent MI and prevalent CHD. For analysis of incident events, participants with a history of MI, CHD or revascularization at the baseline examination were excluded. All-cause mortality after MI or CHD was also investigated with follow-up time from first MI or CHD incident events until death, loss to follow-up, or the end of study.

Statistical analysis

Single variant and gene-based analyses were conducted in each participating cohort respectively, followed by meta-analysis performed for each cardiovascular outcome to summarize results. All autosomal variants were coded to the minor allele observed in the CHARGE jointly called data (21) and assumed log-additive genetic effect in the analyses. The minor allele frequency (MAF) thresholds were defined using the European allele frequencies derived from the CHARGE jointly called data. (21) Variant annotation was performed centrally within CHARGE using dbNSFP. (23, 24) Variants with MAF ≥ 1% were included in single variant tests for prevalent MI and CHD and for incident MI. Single variant results for incident CHD followed the same analytic approach and are reported in Morrison et al. (20) and are not reported in detail here. Gene-based tests were evaluated for MI and CHD outcomes: the Sequence Kernel Association Test (SKAT) (25) and a burden test (26). Only functional coding variants (missense, stop-gain, stop-loss, or splice-site changes) with MAF < 5% were aggregated by gene, and we only analyzed genes with a cumulative minor allele count (cMAC) ≥ 5.

For both single variant and gene-based burden tests of prevalent events, we performed Firth's logistic regression model to test the association between each variant and cardiovascular outcome using the "logistf" package in R (27-29) to account for the possible inflated type one error in the rare variant

163 association analysis in a case-cohort study design.(30) Meta-analysis for prevalent events was conducted
164 with METAL (31) and applied the genomic control correction. For the single variant and two gene-based
165 tests of incident events, a Cox proportional hazards regression model implemented in the seqMeta
166 package in R was used to test the association between each variant and the incident event or post-event
167 all-cause mortality. SeqMeta was used both at the study-specific analysis and meta-analysis levels. (32)
168 All study-specific analyses (single variant and gene-based tests) were adjusted for cohort-specific design
169 variables (e.g. study sites, family structure) and for population substructure using principal components as
170 needed. We applied a Bonferroni corrected threshold to determine statistical significance in each analysis
171 as described below.
172

173 # Results

174 ## Prevalent MI and CHD association

175 A total of 27,349 participants of European ancestry from seven cohorts including 1831 prevalent
176 MI cases (6.7%) and 2518 prevalent CHD cases (9.2%) were used in the meta-analyses of prevalent
177 events (**S1 Table**). We examined individually a total of 36,406 variants, combining both low-frequency
178 and common variants (MAF ≥ 1%), across all autosomal chromosomes corresponding to a Bonferroni
179 corrected significance threshold of P=1.37 × 10^{-6}. A low-frequency (MAF=1.64%) intronic variant
180 (rs988583) in the phospholipase C like 1 gene (*PLCL1*) was significantly associated with prevalent MI
181 (P=7.12 × 10^{-7}; OR=1.80, 95% confidence interval=1.43 to 2.27; **Table 1**). Three common variants were
182 significantly associated with prevalent CHD: rs9349379 in *PHACTR1* and rs1333048 and rs4977574 in
183 the 9p21 region (**Table 1**).

184 **Table 1. Low-frequency and common variants associated with prevalent MI and CHD.**

Outcome	Variant	Chromosome and Position*	Allele 1 / Allele 2	Locus	Function	Frequency of Allele 2 (%)	Odds Ratio (95% Confidence Interval)	p-value

MI	rs988583	2:198987935	C/A	*PLCL1*	Intronic	1.64	1.80 (1.43, 2.27)	7.12×10^{-7}
CHD	rs9349379	6:12903957	A/G	*PHACTR1*	Intronic	39.99	0.84 (0.79, 0.89)	2.96×10^{-8}
CHD	rs1333048	9:22125347	A/C	9p21	Intergenic	48.98	0.85 (0.80, 0.91)	3.25×10^{-7}
CHD	rs4977574	9:22098574	A/G	9p21	Intronic	48.24	0.86 (0.87, 0.91)	1.03×10^{-6}

*Chromosome and nucleotide positions are based on genome build GRCh37.

In the gene-based burden tests, we analyzed 16,628 autosomal genes that contained functional low-frequency or rare variants with MAF < 5% and with a cumulative minor allele count (cMAC) ≥ 5; therefore, the Bonferroni corrected p-value threshold was $P=3.01 \times 10^{-6}$. The transmembrane serine protease 5 gene (*TMPRSS5*) on chromosome 11, containing nine nonsynonymous rare variants (**S2 Table**), was significantly associated with prevalent MI ($P=2.59 \times 10^{-6}$, OR=3.00, 95% confidence interval: 1.90, 4.73; **Table 2**). The low-density lipoprotein receptor class A domain containing 1 gene (*LDLRAD1*) on chromosome 1 contained seven rare variants (**S2 Table**) and was significantly associated with prevalent CHD ($P=1.30 \times 10^{-6}$, OR=4.48, 95% confidence interval: 2.44, 8.23; **Table 2**).

Table 2. Genes associated with prevalent MI and CHD in gene-based analysis.

Outcome	Gene	Chromosome and Position*	cMAC**	Variants (n)^	Test	Odds Ratio (95% Confidence Interval)	p-value
MI	*TMPRSS5*	11:113558268-113577151	152.02	9	Burden	3.00 (1.90, 4.73)	2.59×10^{-6}
CHD	*LDLRAD1*	1:54472971-54483859	60.05	7	Burden	4.48 (2.44, 8.23)	1.30×10^{-6}

*Chromosome and nucleotide positions are based on genome build GRCh37.
**cMAC = overall cumulative minor allele count.
^Variants (n) = number of variants included in the analysis; variants were restricted to those with MAF < 5% and annotated as nonsynonymous, splice-site, or stop loss/gain function.

Incident MI and CHD association

Nine cohorts contributed a total of 55,736 participants of European ancestry to the analyses of incident events, where 3,031 incident MI cases (5.4%) were reported during an average of 15.0 years of

204 follow-up and 5,425 incident CHD cases (9.73%) were reported during an average of 15.6 years of
205 follow-up (**S3 Table**). A total of 37,109 low frequency and common autosomal variants (MAF ≥ 1%)
206 were individually tested for association with incident MI, with adjustment of age, sex, and population
207 substructure. The Bonferroni corrected p-value threshold for single variant analysis of incident MI was
208 P=1.35 × 10^{-6}. Four common variants in the noncoding region at the 9p21 locus were significantly
209 associated with incident MI (**Table 3**). As previously stated, single variant results for incident CHD are
210 reported in Morrison et al. (20) and are not reported here.

211 **Table 3. Low-frequency and common variants associated with incident MI.**

Outcome	Variant	Chromosome and Position*	Allele 1 / Allele 2	Locus	Function	Frequency of Allele 2 (%)	Odds Ratio (95% Confidence Interval)	p-value
MI	rs4977574	9:22098574	A/G	9p21	Intronic	48.55	1.13 (1.08, 1.19)	1.21×10^{-6}
MI	rs10757278	9:22124477	A/G	9p21	Intergenic	48.00	1.16 (1.10, 1.22)	6.61×10^{-9}
MI	rs1333049	9:22125503	G/C	9p21	Intergenic	48.06	1.16 (1.11, 1.22)	5.92×10^{-9}
MI	rs1333048	9:22125347	A/C	9p21	Intergenic	49.13	1.15 (1.09, 1.21)	7.27×10^{-8}

212 *Chromosome and nucleotide positions are based on genome build GRCh37.

213 For the gene-based analyses, we examined 17,574 genes across all autosomal chromosomes for
214 association with incident MI, and the Bonferroni corrected significance level was P=2.85 × 10^{-6}. The ring
215 finger and CCCH-Type domains 2 gene (*RC3H2*) on chromosome 9 was significantly associated with
216 incident MI in the burden test (P=2.99 × 10^{-6}, OR=0.35, 95% confidence interval=0.23, 0.55; **Table 4**)
217 and contained 12 nonsynonymous and one splice-site rare variants (**S4 Table**). No genes were
218 significantly associated with incident MI using SKAT. For the gene-based analyses of incident CHD,
219 16,620 genes were evaluated and the Bonferroni significance levels was P=3.01 × 10^{-6}. Angiopoietin-
220 like 4 (*ANGPTL4*) on chromosome 19 was significantly associated with incident CHD using SKAT
221 (P=1.29 × 10^{-6}; **Table 4**) and contained 10 variants (**S4 Table**), and no gene was significantly associated
222 using the burden test.

223 **Table 4. Genes associated with incident MI and CHD in gene-based analysis.**

Outcome	Gene	Chromosome and Position*	cMAC**	Variants (n)^	Test	Odds Ratio (95% Confidence Interval)	p-value
MI	*RC3H2*	9:125606835-125667562	356.02	13	Burden	0.35 (0.23, 0.55)	2.99×10^{-6}
CHD	*ANGPTL4*	19:8429011-843925	2830.07	10	SKAT	-	1.29×10^{-6}

224 *Chromosome and nucleotide positions are based on genome build GRCh37.
225 **cMAC = overall cumulative minor allele count.
226 ^Variants (n) = number of variants included in the analysis; variants were restricted to those with MAF <
227 5% and annotated as nonsynonymous, splice-site, or stop loss/gain function.
228

229 # Post MI and CHD mortality analysis

230 Among the 3,751 MI and CHD cases from six cohorts that contributed to the analysis of all-cause
231 mortality, there were 1,860 all-cause deaths over a mean 10.9 years of follow-up (**S5 Table**). We
232 examined 36,685 autosomal variants with MAF ≥ 1% in the single variant analysis (Bonferroni corrected
233 significant level of P=1.36 × 10^{-6}) and 17,574 genes in the gene-based analysis (Bonferroni corrected
234 significant level of P=2.85 × 10^{-6}). No single variant or gene reached the significance threshold in the
235 analysis of all-cause mortality among survivors of MI or CHD. We examined the significant variants and
236 genes reported in Tables 1-4 for their relationship with mortality following a MI or CHD event (**S6**
237 **Table**). While these loci were significantly associated with prevalent and incident MI or CHD events,
238 only the 9p21 common variants were nominally associated with all-cause mortality (P < 0.05). The 9p21
239 variants that were associated with reduced risk of prevalent CHD (rs1333048 and rs4977574; **Table 1**),
240 and with increased risk of incident MI (rs1333048, rs4977574, rs10757278, and rs1333049; **Table 3**)
241 were all associated with modestly reduced risk of all-cause mortality (**S6 Table**).
242

243 # Discussion

244 Our study evaluated genetic susceptibility to MI and CHD in cross-sectional and prospective
245 settings among individuals of European ancestry. We confirmed several previously reported loci and

246 newly identified one low-frequency variant and three genes harboring low-frequency and rare coding
247 variants that warrant investigation in future studies.

248 Single variant analysis of prevalent cardiovascular outcomes revealed a low-frequency
249 (MAF=1.64%) intronic variant, rs988583, in *PLCL1* significantly associated with increased risk of MI
250 (P=7.12 × 10^{-7}). *In silico* replication was conducted by a look up of rs988583 and its association with
251 prevalent MI in the Myocardial Infarction Genetics and CARDIoGRAM exome chip meta-analysis public
252 release (33), and there was no significant association with MI (P=0.34). A GWAS of MI and coronary
253 artery disease (CAD) in a Saudi Arab population identified an intergenic variant, rs7421388, near *PLCL1*
254 associated with CAD (P = 4.31 × 10^{-6}) and replicated in an independent sample of Saudi Arabs (P = 5.37
255 × 10^{-7}). (34) In another study of an ethnic Arab population, rs1147169 in *PLCL1* was protective against a
256 low level of high density lipoprotein-cholesterol levels (P = 2.87 × 10^{-7}). (35) In individuals of European
257 ancestry, rs988583 and rs1147169 are in linkage equilibrium (R^2= 0.0043). In addition to these studies,
258 *PLCL1* has been implicated in coronary artery aneurysm in Kawasaki disease and *PLCL1* might play a
259 role in the regulation of vascular endothelial cell inflammation via interference with proinflammatory
260 cytokine expression. (36)

261 A burden test aggregating low-frequency and rare coding variants in genes showed a significant
262 positive association between *TMPRSS5* and prevalent MI (P=2.59 × 10^{-6}) and *LDLRAD1* and prevalent
263 CHD (P=1.30 × 10^{-6}), and a significant protective association between *RC3H2* and incident MI (P=2.99 ×
264 10^{-6}). A significant association between *ANGPTL4* and incident CHD was identified using SKAT (P=1.29
265 × 10^{-6}). The relationship between *ANGPTL4* and CHD has been previously reported, with the
266 rs116843064 missense variant playing a major role in reducing lipid levels and risk of CHD. (33, 37)
267 Serine proteases, such as *TMPRSS5*, are known to be involved in many physiological and pathological
268 processes, and *TMPRSS5* has been implicated in impaired hearing function. (38) Little is known about
269 *LDLRAD1*, with most marked gene expression in lung and fallopian tube (39), and a rare variant in this
270 gene has been associated with breast cancer. (40) Roquin-2 is encoded by *RC3H2* and has been shown to
271 play a key role in posttranscriptional regulation of autoimmunity and inflammatory response. (41) Each of

these genes associated with prevalent or incident cardiovascular outcomes has rare and low-frequency variants underlying the gene burden tests (**S2 and S4 Table**). We identified 11 putative driving variants of these gene-based associations (i.e. those with $p<0.05$ in **S2 and S4 Table**; rs201233178, rs200417674, and rs116913282 in *TMPRSS5*; rs150560713, rs202234131, rs142900519, and rs76122098 in *LDLRAD1*; rs201920127, rs144714368, and rs199901510 in *RC3H2*; and rs116843064 in *ANGPTL4*). An *in silico* replication was not possible due to the rare frequency of these coding variants and their absence in the public release of the Myocardial Infarction Genetics and CARDIoGRAM exome chip meta-analysis or the analysis of CAD in the UK Biobank and the UK Biobank and CARDIoGRAMplusC4D meta-analysis (10, 33) However, it is important to note that rs116843064 of *ANGPTL4* is the same variant found in the single variant analysis conducted for incident CHD by Morrison et al., and this gene is likely to be driving the significant association found in the SKAT analysis of incident CHD. (20) It is of interest that the effect sizes of the gene-based tests (**Tables 2 and 4**) are larger than the single variant test effect sizes (**Tables 1 and 3**), supporting the notion that low-frequency and rare variants may have a more substantial impact on disease risk.

Although there was no statistically significant result found for all-cause mortality after MI or CHD, after accounting for multiple testing, the protective direction of effect for many of our mortality results suggests that genetic variants might contribute differently in various stages of disease manifestation. Specifically, our results highlight differences in the direction of effect for common variants at the 9p21 locus associated with decreased risk of prevalent CHD, increased risk of incident MI, and a nominally significant reduced risk of all cause-mortality following a cardiovascular event. Generally, the loci identified for prevalent disease were not the same as those identified for incident disease, as has been observed in previous studies. (9) Indeed, comparison of prevalent and incident findings (**S6 Table**) shows that the single variant (*PLCL1* locus) and gene-based (*TMPRSS5* locus) results for prevalent MI were not significantly associated with incident MI, and the direction of effects were consistent for *PLCL1* but not *TMPRSS5*. Similarly, the significant gene, *LDLRAD1*, identified for increased risk of prevalent CHD was not significantly associated with incident CHD, but the direction of effect was consistent. The *RC3H2*

298 gene, which showed an inverse association with incident MI, was not significantly associated with
299 prevalent MI and it exhibited an opposite direction of effect. A possible explanation for these observed
300 differences is that genetic studies of cardiovascular diseases are usually conducted with the cross-
301 sectional study design, which has the potential to oversample participants with longer post-event survival
302 (42) and the results do not always replicate in the prospective studies for disease onset and vice versa. (9)
303 Given the limited statistical power of our findings for post-event survival, our study supports the need for
304 substantially larger well-phenotyped cohorts to differentiate effects of variants associated with CHD from
305 post-event mortality.

306 An advantage of this study is that within the setting of the CHARGE Consortium we are able to
307 evaluate and make comparisons between cross-sectional and prospective study designs, and to investigate
308 all-cause mortality following cardiovascular events. There are differing, but overlapping, sample sizes
309 across the various study designs: 27,349 participants from seven cohorts for prevalent outcomes, 55,736
310 participants from nine cohorts for incident outcomes, and 3,751 MI and CHD cases from six cohorts that
311 contributed to the analysis of all-cause mortality. These differing sample sizes influence our power to
312 detect associations, and inferences about similarities and differences across study designs could be due to
313 biological differences or differences in sample sizes. This investigation of low-frequency and rare variants
314 was limited to the variants included on the genotyping platforms (HumanHap300 Duo+ and
315 HumanExome BeadChip, v1.0-1.2, Illumina, Inc., San Diego, CA) and was also limited to individuals of
316 European ancestry. Additionally, although the variants on the genotyping platform and included in our
317 gene-based tests were enriched for coding variants predicted to be causal, we cannot attribute causality to
318 the variants or genes with novel associations. A strength of this study is that the quality of rare variant
319 genotype calling was maximized by the joint clustering performed within CHARGE on thousands of
320 samples (21).

321 In conclusion, this study comprehensively evaluated the relationship between autosomal genetic
322 variation and prevalent and incident cardiovascular outcomes in participants of European ancestry in the

323 context of the CHARGE consortium. We confirmed previously reported loci influencing heart disease
324 risk as well as newly identified several loci associated with MI and CHD that warrant future investigation.
325

326 Acknowledgments

327 The Age, Gene, Environment, Susceptibility Study (AGES) study has been funded by NIH contracts N01-
328 AG-1-2100 and HHSN271201200022C, the NIA Intramural Research Program, Hjartavernd (the
329 Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the
330 Icelandic National Bioethics Committee, VSN: 00-063. The researchers are indebted to the participants
331 for their willingness to participate in the study.
332

333 The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds
334 from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health
335 and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I,
336 HHSN268201700003I, HHSN268201700004I and HHSN268201700005I). The authors thank the staff
337 and participants of the ARIC study for their important contributions. Funding support for "Building on
338 GWAS for NHLBI-diseases: the U.S. CHARGE consortium" was provided by the NIH through the
339 American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419).
340

341 Cardiovascular Health Study: This CHS research was supported by NHLBI contracts
342 HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079,
343 N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants
344 U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, and U01HL130114 with
345 additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS).
346 Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A
347 full list of principal CHS investigators and institutions can be found at https://chs-nhlbi.org/.

The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

The Family Heart Study (FamHS) was supported by the grant R01-HL-117078 from the National Heart, Lung, and Blood Institute, and grant R01-DK-089256 from the National Institute of Diabetes and Digestive and Kidney Diseases.

The Framingham Heart Study (FHS) The National Heart, Lung and Blood Institute's Framingham Heart Study is supported by contract N01-HC-25195.

GeneSTAR was supported by grants from the National Institutes of Health/National Heart, Lung and Blood Institute (HL49762, HL59684, HL071025, HL58625, U01 HL72518, HL089474, HL092165, HL099747, K23HL105897, K23HL094747, HL11006, and HL112064), National Institute of Nursing Research (NR0224103, NR008153), National Institute of Neurological Disorders and Stroke (NS062059), and by a grant from the National Center for Research Resources (M01-RR000052) to the Johns Hopkins General Clinical Research Center. Genotyping services were provided through the RS&G Service by the Northwest Genomics Center at the University of Washington, Department of Genome Sciences, under U.S. Federal Government contract number HHSN268201100037C from the National Heart, Lung, and Blood Institute.

MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491.

374 Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Provision of exome
375 chip genotyping was provided in part by support of NHLBI contract N02-HL-64278 and UCLA CTSI
376 UL1-TR001881, and the S.Calif DRC DK063491.
377
378 From the Rotterdam Study, the authors are grateful to the study participants, the staff, and the
379 participating general practitioners and pharmacists. This work was supported by the Erasmus Medical
380 Center and Erasmus University, Rotterdam; The Netherlands Organisation for the Health Research and
381 Development (ZonMw); the Research Institute for Diseases in the Elderly (014-93-015, RIDE2); the
382 Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the European
383 Commission (DG XII); the Municipality of Rotterdam; The Netherlands Organisation of Scientific
384 Research (NWO) (175.010.2005.011, 911-03-012); the Netherlands Genomics Initiative (NGI) (NWO
385 050-060-810), the Netherlands Organisation for Scientific Research (NWO) (veni 916.12.154).
386
387 The Study of Health in Pomerania (SHIP) and SHIP-TREND both represent population-based studies.
388 SHIP is supported by the German Federal Ministry of Education and Research (Bundesministerium für
389 Bildung und Forschung (BMBF); grants 01ZZ9603, 01ZZ0103, and 01ZZ0403) and the German
390 Research Foundation (Deutsche Forschungsgemeinschaft (DFG); grant GR 1912/5-1). SHIP and SHIP-
391 TREND are part of the Community Medicine Research net (CMR) of the Ernst-Moritz-Arndt University
392 Greifswald (EMAU) which is funded by the BMBF as well as the Ministry for Education, Science and
393 Culture and the Ministry of Labor, Equal Opportunities, and Social Affairs of the Federal State of
394 Mecklenburg-West Pomerania. The CMR encompasses several research projects that share data from
395 SHIP. The EMAU is a member of the Center of Knowledge Interchange (CKI) program of the Siemens
396 AG. SNP typing of SHIP and SHIP-TREND using the Illumina Infinium HumanExome BeadChip
397 (version v1.0) was supported by the BMBF (grant 03Z1CN22). We thank all SHIP and SHIP-TREND
398 participants and staff members as well as the genotyping staff involved in the generation of the SNP data.
399

The Women's Genome Health Study (WGHS) is supported by the National Heart, Lung, and Blood Institute (HL043851, HL080467, HL099355) and the National Cancer Institute (CA047988 and UM1CA182913), with collaborative scientific support and funding for genotyping provided by Amgen.

References

1. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association. Circulation. 2019;139(10):e56-e528.
2. Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. American Heart Journal. 1991;121(1, Part 2):293-8.
3. Swerdlow DI, Holmes MV, Harrison S, Humphries SE. The genetics of coronary heart disease. British Medical Bulletin. 2012;102(1):59-77.
4. O'Donnell CJ, Nabel EG. Genomics of Cardiovascular Disease. New England Journal of Medicine. 2011;365(22):2098-109.
5. Lloyd-Jones DM, Nam BH, D'Agostino RB, Levy D, Murabito JM, Wang TJ, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. JAMA. 2004;291(18):2204-11.
6. Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic Susceptibility to Death from Coronary Heart Disease in a Study of Twins. New England Journal of Medicine. 1994;330(15):1041-6.
7. Murabito JM, Pencina MJ, Nam B-H, D'Agostino RB, Wang TJ, Lloyd-Jones D, et al. Sibling Cardiovascular Disease as a Risk Factor for Cardiovascular Disease in Middle-aged Adults. JAMA. 2005;294(24):3117-23.
8. Dai X, Wiernek S, Evans JP, Runge MS. Genetics of coronary artery disease and myocardial infarction. World journal of cardiology. 2016;8(1):1-23.
9. Dehghan A, Bis JC, White CC, Smith AV, Morrison AC, Cupples LA, et al. Genome-Wide Association Study for Incident Myocardial Infarction and Coronary Heart Disease in Prospective Cohort Studies: The CHARGE Consortium. PloS one. 2016;11(3):e0144997-e.
10. van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. Circulation research. 2018;122(3):433-43.
11. Klarin D, Zhu QM, Emdin CA, Chaffin M, Horner S, McMillan BJ, et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. Nature genetics. 2017;49(9):1392-7.
12. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. Nature Genetics. 2015;47:1121.
13. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nature Genetics. 2017;49:1385.
14. Natarajan P, Bis JC, Bielak LF, Cox AJ, Dörr M, Feitosa MF, et al. Multiethnic Exome-Wide Association Study of Subclinical Atherosclerosis. Circulation: Cardiovascular Genetics. 2016;9(6):511-20.
15. Yu B, Pulit Sara L, Hwang S-J, Brody Jennifer A, Amin N, Auer Paul L, et al. Rare Exome Sequence Variants in CLCN6 Reduce Blood Pressure Levels and Hypertension Risk. Circulation: Cardiovascular Genetics. 2016;9(1):64-70.
16. Flannick J, Mercader JM, Fuchsberger C, Udler MS, Mahajan A, Wessel J, et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. Nature. 2019;570(7759):71-6.
17. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, et al. Exome-wide association study of plasma lipids in >300,000 individuals. Nature Genetics. 2017;49:1758.
18. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nature genetics. 2017;49(3):403-15.

452 19. Stein EA, Mellis S, Yancopoulos GD, Stahl N, Logan D, Smith WB, et al. Effect of a Monoclonal
453 Antibody to PCSK9 on LDL Cholesterol. New England Journal of Medicine. 2012;366(12):1108-18.
454 20. Morrison AC, Fu Y, O'Donnell C. Variants in ANGPTL4 and the Risk of Coronary Artery
455 Disease. New England Journal of Medicine. 2016;375(23):2303-6.
456 21. Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD, et al. Best Practices and Joint
457 Calling of the HumanExome BeadChip: The CHARGE Consortium. PLOS ONE. 2013;8(7):e68095.
458 22. Goldstein JI, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M, et al. zCall: a rare variant
459 caller for array-based genotyping: genetics and population analysis. Bioinformatics (Oxford, England).
460 2012;28(19):2543-5.
461 23. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs
462 and their functional predictions. Human mutation. 2011;32(8):894-9.
463 24. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and
464 their functional predictions and annotations. Human mutation. 2013;34(9):E2393-E402.
465 25. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing
466 data with the sequence kernel association test. American journal of human genetics. 2011;89(1):82-93.
467 26. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases:
468 application to analysis of sequence data. American journal of human genetics. 2008;83(3):311-21.
469 27. Conti E, Izaurralde E. Nonsense-mediated mRNA decay: molecular insights and mechanistic
470 variations across species. Current Opinion in Cell Biology. 2005;17(3):316-25.
471 28. Firth D. Bias Reduction of Maximum Likelihood Estimates. Biometrika. 1993;80(1):27-38.
472 29. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Statistics
473 in Medicine. 2002;21(16):2409-19.
474 30. Wang X. Firth logistic regression for rare variant association tests. Frontiers in Genetics.
475 2014;5(187).
476 31. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide
477 association scans. Bioinformatics (Oxford, England). 2010;26(17):2190-1.
478 32. Voorman A, Brody JB, Chen H, Lumley T, Davis B. seqMeta: Meta-Analysis of Region-Based
479 Tests of Rare DNA Variants. R package version 1.6.7. 2017 [Available from: https://CRAN.R-
480 project.org/package=seqMeta.
481 33. Stitziel NO, Stirrups KE, Masca NGD, Erdmann J, Ferrario PG, König IR, et al. Coding Variation
482 in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. The New England journal of
483 medicine. 2016;374(12):1134-44.
484 34. Wakil SM, Ram R, Muiya NP, Mehta M, Andres E, Mazhar N, et al. A genome-wide association
485 study reveals susceptibility loci for myocardial infarction/coronary artery disease in Saudi Arabs.
486 Atherosclerosis. 2016;245:62-70.
487 35. Wakil SM, Ram R, Muiya NP, Andres E, Mazhar N, Hagos S, et al. A common variant
488 association study reveals novel susceptibility loci for low HDL-cholesterol levels in ethnic Arabs. Clinical
489 Genetics. 2016;90(6):518-25.
490 36. Lin Y-J, Chang J-S, Liu X, Tsang H, Chien W-K, Chen J-H, et al. Genetic variants in
491 PLCB4/PLCB1 as susceptibility loci for coronary artery aneurysm formation in Kawasaki disease in Han
492 Chinese in Taiwan. Scientific reports. 2015;5:14762-.
493 37. Folsom AR, Peacock JM, Demerath E, Boerwinkle E. Variation in ANGPTL4 and risk of
494 coronary heart disease: the Atherosclerosis Risk in Communities Study. Metabolism: clinical and
495 experimental. 2008;57(11):1591-6.
496 38. Guipponi M, Toh M-Y, Tan J, Park D, Hanson K, Ballana E, et al. An integrated genetic and
497 functional analysis of the role of type II transmembrane serine proteases (TMPRSSs) in hearing loss.
498 Human Mutation. 2008;29(1):130-41.
499 39. Gene expression for LDLRAD1 (ENSG00000203985.10) [Internet]. Available from:
500 https://gtexportal.org/home/gene/LDLRAD1.

501 40. Haiman CA, Han Y, Feng Y, Xia L, Hsu C, Sheng X, et al. Genome-wide testing of putative
502 functional exonic variants in relationship with breast and prostate cancer risk in a multiethnic population.
503 PLoS genetics. 2013;9(3):e1003419-e.
504 41. Pratama A, Ramiscal Roybel R, Silva Diego G, Das Souvik K, Athanasopoulos V, Fitch J, et al.
505 Roquin-2 Shares Functions with Its Paralog Roquin-1 in the Repression of mRNAs Controlling T
506 Follicular Helper Cells and Systemic Inflammation. Immunity. 2013;38(4):669-80.
507 42. Anderson CD, Nalls MA, Biffi A, Rost NS, Greenberg SM, Singleton AB, et al. The effect of
508 survival bias on case-control genetic association studies of highly lethal diseases. Circulation
509 Cardiovascular genetics. 2011;4(2):188-96.

510

511 # Supporting information

512 **S1 Document. Characteristics of the participating cohorts.**

513 **S1 Table. Study participants' characteristics for prevalent MI and CHD analysis.**

514 **S2 Table. Low-frequency and rare variants underlying top signals from gene-based analysis for**

515 **prevalent events.**

516 **S3 Table. Study participants' characteristics for incident MI and CHD analysis.**

517 **S4 Table. Low-frequency and rare variants underlying top signals from gene-based analysis for**

518 **incident events.**

519 **S5 Table. Study participants' characteristics for post MI and CHD mortality analysis.**

520 **S6 Table. Prevalent and incident findings in relation to corresponding outcomes.**