

1 **Simultaneous assessment of genetic and epigenetic contributions to plasma lipid levels with**
2 **respect to cardiovascular risk**

3
4 Fumihiko Takeuchi, PhD^{1,2}; Masaya Yamamoto, MD³; Masahiro Nakatochi, PhD⁴; Kozue Takano, PhD
5 ^{2,5}; Atsuko Okazaki, MD⁵; Sakurako Emoto, MD³; Yasuharu Tabara, PhD⁶; Tomohiro Katsuya, MD^{7,8};
6 Ken Yamamoto, MD⁹; Masato Isono, PhD¹; Kotaro Mori, PhD²; Tatsuaki Matsubara, MD¹⁰; Sahoko
7 Ichihara, MD¹¹; Mitsuhiro Yokota, MD¹²; Hisao Hara, MD³; Yukio Hiroi, MD³; Norihiro Kato, MD^{1,2,5*}

8
9 ¹Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global
10 Health and Medicine

11 ²Medical Genomics Center, Research Institute, National Center for Global Health and Medicine

12 ³Department of Cardiology, Center Hospital, National Center for Global Health and Medicine

13 ⁴Public Health Informatics Unit, Nagoya University Graduate School of Medicine

14 ⁵Department of Genomic Medicine, Center Hospital, National Center for Global Health and Medicine

15 ⁶Graduate School of Public Health, Shizuoka Graduate University of Public Health

16 ⁷Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine

17 ⁸Department of Geriatric Medicine and Nephrology, Osaka University Graduate School of Medicine

18 ⁹Department of Medical Biochemistry, Kurume University School of Medicine

19 ¹⁰Faculty of Human Sciences, Aichi Mizuho College

20 ¹¹Department of Environmental and Preventive Medicine, Jichi Medical University School of
21 Medicine

22 ¹²Kurume University School of Medicine

23
24 **Short title:** Genetic and epigenetic risk prediction of lipids

25
26 * Corresponding author

27 Norihiro Kato

28 1-21-1 Toyama, Shinjuku-ku, Tokyo, 162-8655, JAPAN

29 Email: nokato@ri.ncgm.go.jp

30
31 Total word count of the manuscript: 7,067 words (in main text)

32 **Abstract (limited to 250 words)**

33

34 **Background:** This study aims to develop a model for simultaneously assessing genetic and epigenetic
35 contributions to plasma lipid levels.

36 **Methods:** The predictive model was developed using two cardiovascular risk groups, i.e., individuals
37 with high low-density lipoprotein cholesterol (LDL-C) levels (≥ 160 mg/dl, $N = 296$) and coronary artery
38 disease (CAD) ($N = 315$), in contrast to reference (max $N = 3,801$) and non-CAD individuals ($N = 164$).
39 For genetic predisposition, rare pathological variants in five target genes related to familial
40 hypercholesterolemia (FH) were screened, while common variants were characterized to calculate a
41 polygenic risk score (PRS). The methylation risk score (MRS) was also calculated for epigenetic profiles
42 based on DNA methylation levels at 13 CpG sites. A relationship between these variables and lipid
43 levels was analyzed in regression and quantile models.

44 **Results:** A total of 17 rare FH-related gene variants were identified in patients with high LDL-C or CAD,
45 significantly more prevalent than in the general Japanese population (2.8% vs. 0.2%, $P < 1 \times 10^{-15}$). For
46 the rare variants plus PRS, the predictability of individual LDL-C increased (correlation coefficient
47 between predicted and measured values, $r = 0.261$, $P = 1.7 \times 10^{-11}$) compared to PRS alone ($r = 0.151$,
48 $P = 1.2 \times 10^{-4}$). PRS and MRS had the most significant impact on high-density lipoprotein cholesterol
49 and triglycerides, respectively. The two risk scores had additive effects on these traits.

50 **Conclusions:** Our results provide proof-of-concept that assessing the relative contribution of genetic
51 predisposition and DNA methylation levels (reflecting past environmental exposures) may help
52 individuals refine their dyslipidemia treatment.

53

54 **Key Words:** Genetic risk score; Plasma lipids; Coronary artery disease; DNA methylation

55 Dyslipidemia is a significant risk factor for atherosclerotic cardiovascular diseases, including
56 coronary artery disease (CAD), to which genetic, environmental, and demographic factors contribute
57 interactively. Epidemiological studies have shown that appropriate management of dyslipidemia can
58 significantly decrease cardiovascular morbidity and mortality. Atherosclerosis treatment has often
59 focused on reducing low-density lipoprotein cholesterol (LDL-C) among dyslipidemias, with lifestyle
60 modifications emphasized in all patients with high LDL-C.^{1,2} Above all, dietary intervention is essential
61 because high saturated fat intake causes increased concentrations of LDL-C.³ However, sustaining
62 lifestyle modifications over time is generally challenging, so drug therapy is sometimes necessary to
63 reduce the risk of atherosclerosis.¹

64 People with a vital genetic component, e.g., familial hypercholesterolemia (FH) patients, also
65 require lipid-lowering medications such as statins at an earlier stage.² For CAD, while the introduction
66 of statins has considerably changed its management and prevention, the disease burden remains
67 high. Clinical practice guidelines emphasize the importance of estimating the absolute risk of CAD
68 and tailoring the intensity of preventive actions accordingly.^{1,4} The optimal LDL-C goal for persons at
69 low risk is set at <116 mg/dl by the 2019 ESC Guidelines, which used to be set at <160 mg/dl (for
70 persons with 0-1 risk factor) by the NCEP-ATPIII Guidelines. Still, there is a significant gap between
71 guideline-recommended treatment and daily clinical practice of dyslipidemia.⁵

72 An individual's lipid levels are influenced by the degree of environmental exposure and genetic
73 predisposition, although no method has been established to provide quantitative indicators of these
74 combinations. On the one hand, it is difficult to quantitatively assess the impact of unhealthy lifestyle
75 habits on lipid levels. Even so, exposure to some external factors, such as smoking and high plasma
76 lipids, can cause long-lasting DNA methylation changes in blood cells.⁶⁻⁸ These changes can reflect an
77 individual's environmental history, and consequently, lifestyle habits may be evaluated through
78 epigenetic biomarkers.⁹ On the other hand, genetic information as an auxiliary diagnostic tool for
79 dyslipidemia is still in the realm of research, with a few exceptions, such as the genetic diagnosis of
80 FH.² In this line, polygenic risk score (PRS) has recently become popular to summarize the cumulative
81 effects of genetic loci, successfully identified by genome-wide association studies (GWAS) of lipid
82 traits.^{8,10} Nonetheless, there remains considerable uncertainty in individual PRS estimation for PRS-

83 based risk stratification¹¹ and substantial difficulty in unbiased assessment of the contribution of rare
84 variants to complex traits.¹²

85 In light of this situation, we performed genetic (i.e., GWAS array genotyping and rare variant
86 search in FH-related target genes) and epigenetic (specifically, DNA methylation) analyses focusing on
87 lipid traits, specifically LDL-C. We aim to develop a model for evaluating genetic and epigenetic
88 contributions to plasma lipid levels with a risk prediction index simultaneously incorporating multiple
89 data modalities for individuals with high LDL-C or CAD. We studied Japanese populations to overcome
90 challenges in applying PRS and methylation profiles to non-European ancestry groups.^{13,14} We
91 collected individual-level data to test our risk prediction index's accuracy. The study demonstrates
92 that a combination of genetic predisposition and DNA methylation has an additive effect on an
93 individual's lipid levels among people at high risk of atherosclerosis and supports the potential utility
94 of multi-omics data in refining an individual's dyslipidemia treatment.

95

96 **Materials and Methods**

97 Please see the Data Supplement for details about the Materials and Methods.

98 **Study population:** The institutional ethics review board approved this study, and participants gave
99 their written informed consent. The procedures followed the ethical standards of the institutional
100 committee on human experimentation at the National Center for Global Health and Medicine
101 (NCGM).

102 In NCGM, adult patients of Japanese descent were recruited via two separate projects: BIO-CVD
103 and NCGM Biobank, whose enrollment started in 2014 and 2012, respectively. They were categorized
104 into three subgroups based on their plasma LDL-C levels and CAD status (see Supplementary
105 Methods): high LDL-C without CAD, CAD, and non-CAD subgroups (as shown in Table 1). A hospital-
106 based study called BIO-CVD assessed CAD in 475 participants (306 CAD and 169 non-CAD). Samples
107 from the NCGM Biobank were also examined, specifically those with elevated plasma LDL-C (≥ 160
108 mg/dl) ($N = 300$). These samples were used for genetic and DNA methylation analyses (Fig. 1). Also,
109 a significant number of Japanese individuals from non-targeted populations were included. We used
110 publicly available whole genome sequencing data ($N = 8.3K$) for low-frequency (or rare) genetic

111 variants. For common single nucleotide variants (SNVs), we used GWAS array genotyping data in
112 three population-based cohorts (including the Kita-Nagoya Genomic Epidemiology [KING] study
113 cohort¹⁵) and another hospital-based sample at NCGM (named NCGM hospital cohort; $N = 351$)
114 (Table S1), which we referred to hereafter as reference individuals in total ($N = 3,801$). Furthermore,
115 for DNA methylation, we utilized epigenome-wide association study (EWAS)-array assayed data for
116 part ($N = 314$) of the KING study cohort samples previously reported.¹⁵

117 When analyzing prediction models for dyslipidemia, the high LDL-C and CAD subgroups were set
118 as atherosclerotic risk groups, while the KING study cohort samples and the non-CAD subgroup were
119 set as comparison groups (Fig. 1). With these subgroup analyses, we attempted to test whether high
120 LDL-C or CAD status affects the association between DNA methylation and lipid traits.

121 Medical records were reviewed to extract the participants' plasma lipid levels and medication
122 information at enrollment and clinical conditions of CAD patients (Table 1). To ensure the reliability of
123 the data, we excluded individuals who had missing information on key variables: plasma lipid levels,
124 genotype, and DNA methylation data.

125 **Imputation of baseline LDL-C level:** For the imputation of baseline (or pretreatment) LDL-C, we
126 performed a retrospective analysis on data from 36 individuals with high LDL-C in whom plasma LDL-
127 C measurements were available both before and within 18 months after initiating statins and
128 ezetimibe (Fig. S1). Similar to the study by Ruel *et al.*,¹⁶ we performed imputation of baseline LDL-C
129 by dividing the on-treatment value by the proportion obtained by subtracting the expected
130 treatment-induced reduction (expressed as a ratio) from 1.

131 **Gene sequencing:** Next-generation sequencing screened mutations in five FH-related genes (*LDLR*,
132 *APOB*, *PCSK9*, *LDLRAP1*, and *APOE*). The sequences were aligned to a human genome reference
133 sequence (GRCh37/hg19). SNVs and insertion/deletion (Indel) were called using the software GATK
134 (<https://gatk.broadinstitute.org/>) and McCortex (<https://github.com/mcveanlab/mccortex>) and
135 annotated using VarSeq (<https://www.goldenhelix.com/products/VarSeq/>). We then retrieved rare,
136 putatively pathogenic variants, including SNVs that cause non-synonymous, nonsense, or splice-site
137 substitutions and are predicted to be deleterious by ClinVar, LOVD, or *in silico* prediction algorithms
138 in dbNSFP.¹⁷

139 **Genotyping:** 767 samples were newly genotyped with the Infinium OmniExpress-24 BeadChip
140 (Illumina, San Diego, CA, USA). Besides, GWAS array genotyping data for reference individuals¹⁸ were
141 utilized.

142 **Polygenic risk score (PRS):** For each lipid trait, a PRS was built using LDpred
143 (<https://github.com/bvilhjal/ldpred>)¹⁹ and external GWAS summary statistics,²⁰ for which potential
144 population stratification was adjusted using genetic principal components.²¹ Briefly, for each linkage
145 disequilibrium (LD)-based clump across the genome, index SNPs passing specific statistical
146 significance ($P < 0.005$) and LD coefficient ($r^2 \geq 0.5$) thresholds were extracted from large GWAS
147 reported in Japanese.²⁰ Individuals were scored based on the number of risk alleles (0, 1, or 2)
148 weighted by beta of the SNP-trait association, thereby used for developing the best polygenic
149 predictive indicator of LDL-C (PRS_{LDL-C}) and other lipid traits.

150 **Conversion of PRS to LDL-C value:** We developed a formula that converts PRSs to LDL-C values using
151 two different external reference panels to estimate the standardized effects of a PRS_{LDL-C} on plasma
152 LDL-C level. First, we set standardized PRS_{LDL-C} decile classes in the general Japanese population and
153 used them as the Japanese standard PRS for LDL-C (Fig. S2A-C). Then, we calculated the mean LDL-C
154 values for Japanese standard PRS decile classes in the NCGM hospital cohort, which was not
155 previously screened for cardiovascular disease (Table S1 and Fig. S3). The mean values of LDL-C for
156 each PRS decile class (shown in Fig. S2D) were used to convert PRS_{LDL-C} in individuals from BIO-CVD
157 and NCGM Biobank (Fig. 3D and Fig. S2E) belonging to the same population as the NCGM hospital
158 cohort.

159 **Genetic composite risk:** For rare [minor allele frequency (MAF) < 0.01] variants of putative functional
160 significance, we arbitrarily classified them into three distinct categories—(1) disruptive (frameshift or
161 splice-donor), (2) damaging, and (3) all other non-synonymous variants—similar to the previous
162 study²² with some modifications (see the Supplementary Methods). We estimated the net increase
163 in LDL-C levels for rare variant carriers by subtracting a predicted value (based on PRS_{LDL-C}) from the
164 actual measured value. Using linear regression models, we calculated LDL-C changes from individuals
165 without rare FH-related gene variants and the standard error (SE) by category, including disruptive,
166 damaging, and other non-synonymous variants.

167 To combine the genetic risk of rare variants and SNVs, we used the following algorithm: [an
168 individual's LDL-C value] = [predicted LDL-C value based on PRS_{LDL-C}] + [estimated changes from non-
169 carriers (by rare variant category)].

170 **Methylation risk score (MRS):** Genomic DNA was extracted from the peripheral blood buffy coat and
171 stored at -80°C for DNA methylation analysis. As previously reported,²³ genome-wide methylation
172 profiling was performed with the Illumina EPIC array (Illumina Inc. San Diego, CA, U.S.A.). Also, droplet
173 digital PCR (ddPCR) assay in the QX200 Droplet Digital PCR system (Bio-Rad, Carlsbad, CA, U.S.A.) was
174 employed to measure methylation extent at 13 selected CpG sites, which had been reported to show
175 robust evidence for association with lipid traits in previous EWASs (Supplementary Methods, Table
176 S3).¹⁰

177 Because methylation profiles reflect environmental effects, MRSs were created for lipid traits
178 using weighted sums of beta values from 13 CpG sites.²⁴ Here, the effect sizes from the previous
179 multi-ancestry EWAS for lipid traits¹⁰ were utilized as external weights for each CpG probe.

180 **Statistical analysis:** Pearson's correlation coefficient, r , was calculated to measure the strength and
181 direction of the linear relationship between actual lipid levels and those predicted by the individual
182 risk score models. To develop a prediction model, the relationship between predictor variables (i.e.,
183 PRS, MRS, and/or PRS+MRS) and outcome variables (i.e., plasma lipid levels) was examined in this
184 study by two analytical methods: quantile binning and multivariable regression analysis. Quantile
185 binning was used to convert an individual's PRS to the absolute scale (i.e., plasma lipid value) as
186 previously reported.²⁵ In situations where the correctness of the linearity assumption is
187 undetermined, quantile binning is essential to correct LDL-C changes due to rare functional variants
188 and to combine them with the MRS at the individual level in a representative sample. Multivariable
189 regression analyses, on the other hand, were performed to quantify the predictive utility of predictors
190 (i.e., risk scores) at the group level separately by (sub)group, and the coefficient of determination
191 (denoted R^2) was used to provide a measure of how well the observed outcomes are approximated
192 by the regression predictors in the tested population. The imputed value for lipid-lowering therapy
193 was used in analyses including PRS when analyzing LDL-C.

194 No formal sample size calculations were performed since both BIO-CVD and NCGM Biobank are

195 ongoing studies and there is no generally accepted approach to estimating the sample size needed
196 to develop a prediction model like this study.

197 Significant results are shown as mean \pm SE with $P < 0.05$ unless otherwise stated.

198

199 **Results**

200 ***Imputation of baseline LDL-C level***

201 This study recruited adult patients with high LDL-C and CAD at NCGM; of 315 CAD patients, 157 had
202 a clear history of myocardial infarction, but 158 did not (Table 1). A considerable part of the
203 participants had been prescribed statins, with 19% in the high LDL-C subgroup and 70% in the CAD
204 subgroup (Table 1). For statin, irrespective of its type and dose, the average percent reduction in LDL-
205 C from baseline (or pre-treatment) was approximately 40%. Also, for ezetimibe, the percent reduction
206 in LDL-C was 22% (observation in a single patient) (Fig. S1). Hence, we used these values (40% for
207 statins and 20% for ezetimibe) to impute pre-treatment LDL-C measurements for patients on lipid-
208 lowering therapy.

209 ***Rare variants in five FH-related genes***

210 We identified 1,762 variants in five FH-related genes, of which 58 were rare (MAF < 0.01) variants with
211 functional significance. The variants (30 in *APOB*, 13 in *LDLR*, 7 in *PCSK9*, 1 in *LDLRAP1*, and 7 in *APOE*)
212 could be classified into three categories—2 disruptive (frameshift and splice-donor each), 13
213 damaging, and 43 other non-synonymous variants (Table 2 and Table S2)—according to criteria
214 described in the Methods. In a CAD patient, we also identified a gain-of-function type variant for
215 *PCSK9* (indel, TGCCAGCGCCT/-), which appeared to be protective against CAD by exerting LDL-C
216 lowering effects (Table 2). None of the identified rare variants overlapped with variants reported to
217 affect response to statin therapy.²⁶

218 We evaluated the effect sizes of rare FH-related gene variants compared to a reference group of
219 non-carriers by category to calculate the net increase in LDL-C caused by individual rare variants (see
220 Methods). LDL-C levels increased significantly ($P < 0.05$) in all three categories compared to the
221 reference group: plus 106, 42, and 13 mg/dl for disruptive, damaging, and non-synonymous rare
222 variants, respectively (Fig. 2A). The order of increase was consistent with the predicted pathogenicity

223 of identified rare variants (Fig. 2B) and also with results from another study in Japanese.²²

224 The ratio of carriers for ‘disruptive + damaging’ variants was significantly ($P < 1 \times 10^{-15}$) higher
225 among CAD and high LDL-C subjects (ratio=0.025 and 0.03, respectively) than in the general
226 population, TMM (ratio=0.002). (Fig. 2C).

227 **PRS and composite genetic risk prediction for LDL-C**

228 By use of the standardized PRS_{LDL-C} decile classes, we found that PRS_{LDL-C} was shifted towards higher
229 (from 7th to 10th) decile classes among high LDL-C subjects, whereas PRS_{LDL-C} distribution was
230 protruded in the 9th decile class among CAD subjects (Fig. 3C).

231 We examined the correlation between predicted and actual LDL-C levels in newly recruited
232 individuals at NCGM, based on standardized PRS_{LDL-C} decile and estimates of LDL-C increase by rare
233 variant category in Japanese. (Fig. 3D,E). The correlation was stronger for composite genetic risk
234 prediction with both PRS_{LDL-C} and rare variant effects (middle plot; $r=0.261$, $P = 1.7 \times 10^{-11}$) than
235 conventional genetic risk prediction with PRS_{LDL-C} alone (left plot; $r=0.151$, $P = 1.2 \times 10^{-4}$).

236 **Impacts of PRSs and MRSs on lipid traits**

237 Each subgroup (Table 1) had unimodal distributions of HDL-C, triglycerides, and BMI, distinct from
238 LDL-C (Fig. S4). Between high LDL-C and CAD subgroups, there were significant differences in BMI
239 ($P=0.004$) and HDL-C ($P=9.2 \times 10^{-18}$) in addition to LDL-C ($P=5.8 \times 10^{-158}$).

240 We confirmed the consistency ($R^2=0.34-0.86$) of two DNA methylation analytical methods
241 (EPIC array and ddPCR; see Fig. S5). Then, we used ddPCR to measure the methylation levels of 13
242 CpGs (see Table S3) and found that the effect sizes of the CpG-trait association were well correlated
243 with previous multi-ethnic EWAS (Fig. 4A). Furthermore, there were significant differences ($P=0.0045$
244 for LDL, $P=3.5 \times 10^{-7}$ for HDL) in the MRS distribution between the high LDL-C and CAD subgroups (Fig.
245 S6), consistent with phenotypic differences between the subgroups.

246 When the CpGs were analyzed individually, the CpG-trait association was more robust in the
247 CAD subgroup compared to the high LDL-C subgroup for lipid traits (Fig. 4A). This was supported by
248 the results of the MRS-trait association, which involves the cumulative effects of top-hit CpGs
249 previously reported for each lipid trait (Fig. 4B and Table S4). Contrarily, the impact of PRSs on lipid
250 traits tended to be smaller in the CAD subgroup than in the high LDL-C subgroup. The two risk scores

251 were also explanatory factors for the variance of lipid traits in the general population (as represented
252 by the KING study cohort). At the same time, there were some differences in their relative influence
253 compared to the high LDL-C and CAD subgroups. Overall, within each subgroup or population, the
254 effects of PRSs and MRSs on lipid traits were additive (Fig. 4B). Although it remains a preliminary
255 finding given the modest sample size in the current non-CAD subgroup, the strength of association
256 between the risk scores and lipid traits appeared to be influenced by CAD status or its related factors
257 (Table S4).

258 ***Predictability of lipid traits by risk score models***

259 We compared correlation strengths between predicted and measured lipid trait values using various
260 risk score models to assess individual predictability (Fig.3D-F and Fig. S7). The model incorporating
261 PRS and MRS (plus rare variants for LDL-C) showed a higher correlation than the one using only one
262 risk score, supporting its efficacy in enhancing predictability.

263

264 **Discussion**

265 This study, focusing on cardiovascular risk, evaluates the relative contribution of genetic
266 predisposition and DNA methylation to plasma lipid levels, by developing a prediction model for
267 dyslipidemia. Our analysis newly demonstrates two key points. First, concerning genetic
268 predisposition, the effect sizes of rare FH-related gene variants are generally more prominent than
269 the integrated effect of common variants (i.e., PRS). Combining the two can increase the accuracy of
270 predicting an individual's genetic predisposition for LDL-C. Second, DNA methylation, which is
271 assumed to reflect an individual's history of environmental exposure, interacts with genetic
272 predisposition to affect an individual's blood lipid levels. Therefore, combining DNA methylation and
273 genetic predisposition can enhance the accuracy of predicting optimal treatment.

274 As a first step in this study, we have successfully developed an algorithm to predict composite
275 genetic risk for high LDL-C levels. Prioritizing the detection of rare pathogenic variants in population
276 screening, as well as the concurrent use of PRS is a topic of active discussion for genomic conditions
277 that are actionable and adult-onset.²⁷ Thus, a study has shown a risk prediction model that combines
278 rare pathogenic variants, PRS, and individual risk factor variables for breast cancer (i.e., dichotomous

279 trait).²⁸ However, this has not been previously known for lipids, a quantitative trait. The heritability of
280 blood lipid levels, based on SNPs, is estimated to be relatively high (23.3% for LDL-C, 30.5% for
281 triglycerides, and 32.8% for HDL-C); however, the susceptibility loci identified so far can only partially
282 explain it.²⁹ Among the lipid traits, causative gene variants with a strong genetic influence, or
283 monogenic mutations are known principally for LDL-C and are applied in current clinical practice for
284 the genetic diagnosis of FH. In a population-based sample of European descent, only 2% of those
285 with severe hypercholesterolemia (LDL-C > 183 mg/dl) were reported to carry a monogenic mutation,
286 and 23% had a high PRS (top 5th percentile) for LDL-C, indicating the importance of assessing both
287 monogenic and polygenic models simultaneously.³⁰ On the other hand, in a hospital-based setting,
288 only about 40% of patients clinically diagnosed with FH were estimated to have an FH-causing rare
289 variant, with the remainder likely to be polygenic in origin.³¹ Genetic effects on LDL-C levels have also
290 been studied in relation to myocardial infarction susceptibility, with similar results to the present
291 study for the FH-related genes *LDLR* and *PCSK9*.²² Accordingly, the newly developed algorithm for
292 predicting LDL-C levels is expected to optimize preventive therapy for CAD patients.

293 In addition to genetic predisposition, epigenetics, specifically DNA methylation, has attracted
294 attention in mediating gene-environment interactions in cardiometabolic phenotypes. This is
295 because nutritional and lifestyle factors considerably influence epigenetics.⁶⁻⁸ Also, EWAS has shown
296 through longitudinal analyses that smoking and alcohol consumption are associated with individual
297 differences in DNA methylation at many CpG sites in the genome,^{6,7} and that such an association
298 between traits and CpG methylation is likely to reflect to a significant degree the reverse causation of
299 cardiometabolic phenotypes, including obesity.³² Furthermore, it has recently been noted that GWAS
300 and EWAS capture different aspects of the biology of complex traits.³³ Thus, interest has been
301 growing in applying the PRS approaches to DNA methylation data, the so-called MRS. Still,
302 methodological challenges must be resolved in its construction and use as a biomarker for
303 environmental exposures.³⁴

304 Given these circumstances, we have evaluated lipid MRSs containing a list of 13-CpGs in
305 combination with PRS or without PRS through the development of prediction models for individual
306 trait values. We initially selected the 13 CpGs because they showed a strong association with any of

307 the lipid traits in the multi-ethnic EWAS,¹⁰ but these CpG-trait associations were reported to extend
308 beyond lipids, considerably overlapping with other cardiometabolic phenotypes such as blood
309 pressure, alcohol intake, and liver fat (Table S5). This leads to a hypothesis that some upstream
310 environmental determinant (e.g., diet or exercise) is responsible for the overlap. For lipids, the effect
311 sizes for single CpG-trait association between the present study in Japanese and previous multi-
312 ethnic EWAS¹⁰ correlated well (Fig. 4A), which corroborates the reported high correlation between
313 Europeans and non-Europeans (e.g., African Americans).¹⁰ Although there are some differences
314 between subgroups in the interindividual variance explained by the MRS, the MRS predicts lipid traits
315 in an additive manner independent of the PRS, particularly prominent among CAD subjects (Figure
316 4B and Table S4). This is in accord with previous findings on BMI, suggesting that the MRS represents
317 environmental effects.²⁴

318 While the clinical benefits of LDL-C lowering are widely acknowledged for the prevention of
319 atherosclerotic cardiovascular diseases, a residual risk beyond LDL-C, including higher plasma levels
320 of triglycerides and triglyceride-related apolipoproteins, has become recognized. In general,
321 management of hypertriglyceridemia starts with lifestyle modification, and the use of drugs may be
322 considered in high-risk individuals with limited benefits from lifestyle modification. This necessitates
323 the encouragement of patient adherence to lifestyle changes or drug regimens.¹

324 We envision that the combination of PRS (or, more preferably, composite genetic risk prediction)
325 and MRS will contribute to a tailored approach to the management of dyslipidemia and even CAD, as
326 emphasized by the clinical practice guidelines.^{1,4} In view of the considerable individual variability in
327 not only dietary habits but also the LDL-C response to dietary and drug treatment,³⁵ the use of our
328 risk prediction index (exemplified in Fig. S8) in clinical practice may aid patient-doctor communication,
329 facilitating adherence to and optimization of lipid treatment.

330 There are strengths and limitations in this study. First, we investigated the impact of genetic and
331 DNA methylation variations on dyslipidemia in two risk groups (high LDL-C and CAD subjects), in
332 contrast to the general population and non-CAD subjects. It is a strength that this information can be
333 used to optimize treatment for primary (high LDL-C) and secondary (CAD) prevention of
334 atherosclerotic cardiovascular disease by considering the relative contributions of genetics and

335 environmental exposure. However, given age differences, individual differences in various lifestyle
336 habits, and the impact of drug therapy on blood lipid measurements, data collected over time
337 (including pretreatment measurements) in larger populations should be used to improve the
338 accuracy of dyslipidemia risk prediction. Second, although we used 13 CpGs selected from top-hit
339 CpGs in the multi-ethnic EWAS of lipids,¹⁰ the method for determining a set of CpGs for the MRS is
340 still largely undefined. To improve the precision and practicality of the prediction model, an increased
341 number of methylation sites for MRS and standard procedures for optimal weighting and clamping
342 of target CpG sites are needed.¹⁴ Third, although the interaction term was included in multivariate
343 regression analysis, potential interactions between MRS and PRS remain to be deeply explored.
344 Fourth, combining rare FH-related gene variants with PRS can improve the predictability of LDL-C. Still,
345 it must be more feasible and cost-effective to achieve similar predictability for HDL-C and triglycerides
346 in clinical practice. Nevertheless, as the cost of sequencing decreases and genetic variation
347 annotation becomes more manageable, a similar approach for the latter two lipid traits will be
348 justifiable. Fifth, we utilized Japanese original reference panels with individual-level data for SNPs and
349 the phenotypic trait of interest to standardize the PRS, which requires an ancestry-matched
350 reference.¹³ For MRS, external weights from the multi-ethnic EWAS¹⁰ were used to calculate it, after
351 confirming a high correlation between ethnic groups. However, further investigation is necessary to
352 improve the generalizability of our risk prediction index. Sixth, given the limited number of CAD
353 patients in the general Japanese population, the current findings need to be further validated and
354 replicated with a larger, more representative population independently.

355 In summary, we conducted an integrative assessment towards precision medicine for
356 dyslipidemia. Our data provided proof-of-concept that an individual's dyslipidemia treatment
357 regimen could be determined more precisely by the relative contribution of genetic predisposition
358 and DNA methylation levels, which reflect past environmental exposures to a significant degree.

359 **Acknowledgments:**

360 We thank the Research Institute, NCGM staff for their technical assistance with DNA preparation and
361 epigenetic analysis.

362

363 **Sources of funding:**

364 This study was supported by a grant (19A2004) from NCGM and AMED under Grant Number
365 JP22ek0210165 and JSPS KAKENHI Grants (JP20K10514, JP21H03206, JP22H03350).

366

367 **Disclosures:** None

368

369 **Data availability:**

370 All data produced in the present study are available upon reasonable request to the authors.

368 **References**

- 369 1. Authors/Task Force Members; ESC Committee for Practice Guidelines (CPG); ESC National
370 Cardiac Societies. 2019 ESC/EAS guidelines for the management of dyslipidaemias: Lipid
371 modification to reduce cardiovascular risk. *Atherosclerosis*. 2019;290:140-205.
- 372 2. Migliara G, Baccolini V, Rosso A, D'Andrea E, Massimi A, Villari P, De Vito C. Familial
373 Hypercholesterolemia: A Systematic Review of Guidelines on Genetic Testing and Patient
374 Management. *Front Public Health*. 2017;5:252
- 375 3. Chiu S, Williams PT, Krauss RM. Effects of a very high saturated fat diet on LDL particles in adults
376 with atherogenic dyslipidemia: A randomized controlled trial. *PLoS One*. 2017;12:e0170664.
- 377 4. Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, Goldberg AC, Gordon
378 D, Levy D, Lloyd-Jones DM, et al. American College of Cardiology/American Heart Association
379 Task Force on Practice Guidelines. 2013 ACC/AHA guideline on the treatment of blood
380 cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American
381 College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*.
382 2014;129:S1-45.
- 383 5. Dixon DL, Sharma G, Sandesara PB, Yang E, Braun LT, Mensah GA, Sperling LS, Deedwania PC,
384 Virani SS. Therapeutic Inertia in Cardiovascular Disease Prevention: Time to Move the Bar. *J Am
385 Coll Cardiol*. 2019;74:1728-1731.
- 386 6. Dugué PA, Jung CH, Joo JE, Wang X, Wong EM, Makalic E, Schmidt DF, Baglietto L, Severi G,
387 Southey MC, et al. Smoking and blood DNA methylation: an epigenome-wide association study
388 and assessment of reversibility. *Epigenetics*. 2020;15:358-368.
- 389 7. Dugué PA, Wilson R, Lehne B, Jayasekara H, Wang X, Jung CH, Joo JE, Makalic E, Schmidt DF,
390 Baglietto L, et al. Alcohol consumption is associated with widespread changes in blood DNA
391 methylation: Analysis of cross-sectional and longitudinal data. *Addict Biol*. 2021;26:e12855
- 392 8. Jones AC, Irvin MR, Claas SA, Arnett DK. Lipid Phenotypes and DNA Methylation: a Review of the
393 Literature. *Curr Atheroscler Rep*. 2021;23:71.
- 394 9. Santaló J, Berdasco M. Ethical implications of epigenetics in the era of personalized medicine.
395 *Clin Epigenetics*. 2022;14:44.
- 396 10. Jhun MA, Mendelson M, Wilson R, Gondalia R, Joehanes R, Salfati E, Zhao X, Braun KVE, Do AN,
397 Hedman ÅK, et al. A multi-ethnic epigenome-wide association study of leukocyte DNA
398 methylation and blood lipids. *Nat Commun*. 2021;12:3987.
- 399 11. Ding Y, Hou K, Burch KS, Lapinska S, Privé F, Vilhjálmsón B, Sankararaman S, Pasaniuc B.
400 Large uncertainty in individual polygenic risk score estimation impacts PRS-based risk

- 401 stratification. *Nat Genet.* 2022;54:30-39.
- 402 12. Evans LM, Romero Vilella PN. How rare mutations contribute to complex traits. *Nature.*
403 2023;614:418-419.
- 404 13. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat*
405 *Rev Genet.* 2018;19:581-590.
- 406 14. Hüls A, Czamara D. Methodological challenges in constructing DNA methylation risk scores.
407 *Epigenetics.* 2020;15:1-11.
- 408 15. Nakatochi M, Ichihara S, Yamamoto K, Ohnaka K, Kato Y, Yokota S, Hirashiki A, Naruse K, Asano
409 H, Izawa H, et al. Epigenome-wide association study suggests that SNPs in the promoter region
410 of RETN influence plasma resistin level via effects on DNA methylation at neighbouring sites.
411 *Diabetologia.* 2015;58:2781-2790.
- 412 16. Ruel I, Aljenedil S, Sadri I, de Varennes É, Hegele RA, Couture P, Bergeron J, Wanneh E, Baass A,
413 Dufour R, et al. Imputation of Baseline LDL Cholesterol Concentration in Patients with Familial
414 Hypercholesterolemia on Statins or Ezetimibe. *Clin Chem.* 2018;64:355-362.
- 415 17. Liu X, Li C, Mou C, Dong Y, Tu Y. dbNSFP v4: a comprehensive database of transcript-specific
416 functional predictions and annotations for human nonsynonymous and splice-site SNVs.
417 *Genome Med.* 2020;12:103.
- 418 18. Takeuchi F, Akiyama M, Matoba N, Katsuya T, Nakatochi M, Tabara Y, Narita A, Saw WY, Moon
419 S, Spracklen CN, et al. Interethnic analyses of blood pressure loci in populations of East Asian
420 and European descent. *Nat Commun.* 2018;9:5052.
- 421 19. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, Genovese G, Loh PR, Bhatia
422 G, Do R, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am*
423 *J Hum Genet.* 2015;97:576-592.
- 424 20. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, Iwata N, Ikegawa S, Hirata
425 M, Matsuda K, et al. Genetic analysis of quantitative traits in the Japanese population links cell
426 types to complex human diseases. *Nat Genet.* 2018;50:390-400.
- 427 21. Sakaue S, Hirata J, Kanai M, Suzuki K, Akiyama M, Lai Too C, Arayssi T, Hammoudeh M, Al Emadi
428 S, Masri BK, et al. Dimensionality reduction reveals fine-scale structure in the Japanese
429 population with consequences for polygenic risk prediction. *Nat Commun.* 2020;11:1569.
- 430 22. Tajima T, Morita H, Ito K, Yamazaki T, Kubo M, Komuro I, Momozawa Y. Blood lipid-related low-
431 frequency variants in LDLR and PCSK9 are associated with onset age and risk of myocardial
432 infarction in Japanese. *Sci Rep.* 2018;8:8107.
- 433 23. Takeuchi F, Takano K, Yamamoto M, Isono M, Miyake W, Mori K, Hara H, Hiroi Y, Kato N. Clinical

- 434 Implication of Smoking-Related Aryl-Hydrocarbon Receptor Repressor (AHRR) Hypomethylation
435 in Japanese Adults. *Circ J*. 2022;86:986-992.
- 436 24. Shah S, Bonder MJ, Marioni RE, Zhu Z, McRae AF, Zhernakova A, Harris SE, Liewald D, Henders
437 AK, Mendelson MM, et al. Improving Phenotypic Prediction by Combining Genetic and
438 Epigenetic Associations. *Am J Hum Genet*. 2015;97:75-85.
- 439 25. Pain O, Gillett AC, Austin JC, Folkersen L, Lewis CM. A tool for translating polygenic scores onto
440 the absolute scale using summary statistics. *Eur J Hum Genet*. 2022;30:339-348.
- 441 26. Yow HY, Hamzah S, Abdul Rahim N, Suppiah V. Pharmacogenomics of response to statin
442 treatment and susceptibility to statin-induced adverse drug reactions in Asians: a scoping review.
443 *Asian Biomed (Res Rev News)*. 2023;17:95-114.
- 444 27. Lacaze P, Manchanda R, Green RC. Prioritizing the detection of rare pathogenic variants in
445 population screening. *Nat Rev Genet*. 2023;24:205-206.
- 446 28. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, Babb de Villiers C, Izquierdo
447 A, Simard J, Schmidt MK, et al. BOADICEA: a comprehensive breast cancer risk prediction model
448 incorporating genetic and nongenetic risk factors. *Genet Med*. 2019;21:1708-1718.
- 449 29. Khera AV, Won HH, Peloso GM, Lawson KS, Bartz TM, Deng X, van Leeuwen EM, Natarajan P,
450 Emdin CA, Bick AG, et al. Diagnostic Yield and Clinical Utility of Sequencing Familial
451 Hypercholesterolemia Genes in Patients With Severe Hypercholesterolemia. *J Am Coll Cardiol*.
452 2016;67:2578-2589.
- 453 30. Natarajan P, Peloso GM, Zekavat SM, Montasser M, Ganna A, Chaffin M, Khera AV, Zhou W,
454 Bloom JM, Engreitz JM, et al. Deep-coverage whole genome sequences and blood lipids among
455 16,324 individuals. *Nat Commun*. 2018;9:3391.
- 456 31. Sharifi M, Futema M, Nair D, Humphries SE. Genetic Architecture of Familial
457 Hypercholesterolaemia. *Curr Cardiol Rep*. 2017;19:44.
- 458 32. van Dijk SJ, Tellam RL, Morrison JL, Muhlhausler BS, Molloy PL. Recent developments on the role
459 of epigenetics in obesity and metabolic disease. *Clin Epigenetics*. 2015;7:66.
- 460 33. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, Tsai PC, Ried JS, Zhang W, Yang Y, et al.
461 Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity.
462 *Nature*. 2017;541:81-86.
- 463 34. Battram T, Gaunt TR, Relton CL, Timpson NJ, Hemani G. A comparison of the genes and genesets
464 identified by GWAS and EWAS of fifteen complex traits. *Nat Commun*. 2022;13:7816.
- 465 35. Boekholdt SM, Hovingh GK, Mora S, Arsenault BJ, Amarenco P, Pedersen TR, LaRosa JC, Waters
466 DD, DeMicco DA, Simes RJ, et al. Very low levels of atherogenic lipoproteins and the risk for

467 cardiovascular events: a meta-analysis of statin trials. *J Am Coll Cardiol.* 2014;64:485-494.

Fig.1

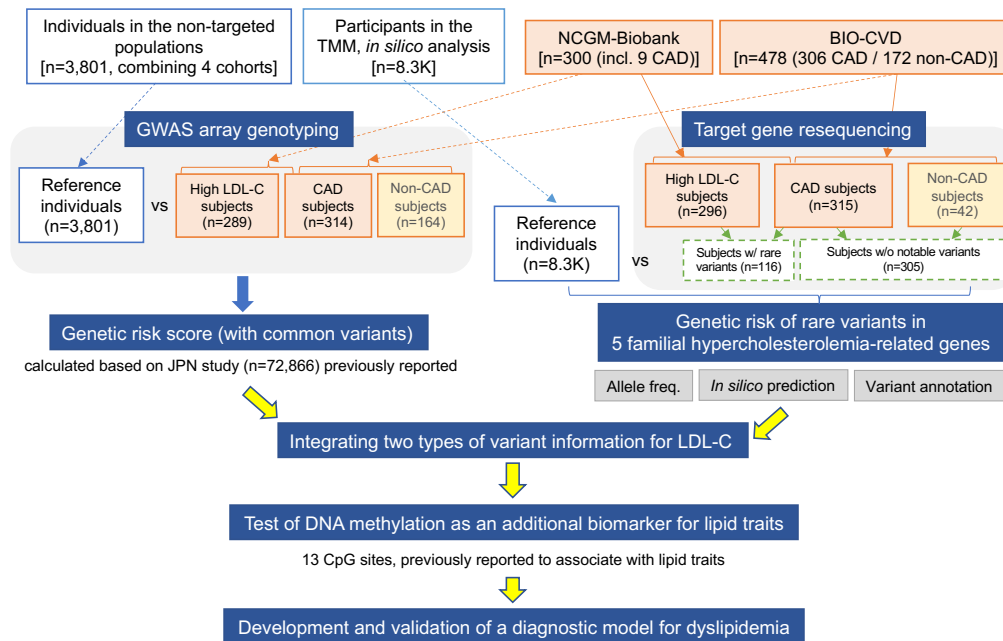


Fig. 1. Overview of study design.

The integrative genomic analysis involves GWAS array genotyping, target gene resequencing, and measuring DNA methylation at selected CpG sites. In contrast to two cardiovascular risk groups—high LDL-C and CAD subgroups, individuals in the non-targeted populations ($N = 3,801$) were used as reference datasets for GWAS array genotyping. The non-CAD subgroup ($N = \sim 164$) was compared to the CAD subgroup by GWAS array genotyping and target gene resequencing. Participants in Tohoku Medical Megabank (TMM) ($N = 8.3K$) were also used as large *in silico* genetic datasets for the evaluation of rare variant frequencies. Moreover, part ($N = 314$) of the KING study cohort was used as a reference dataset for DNA methylation analysis.

Fig.2

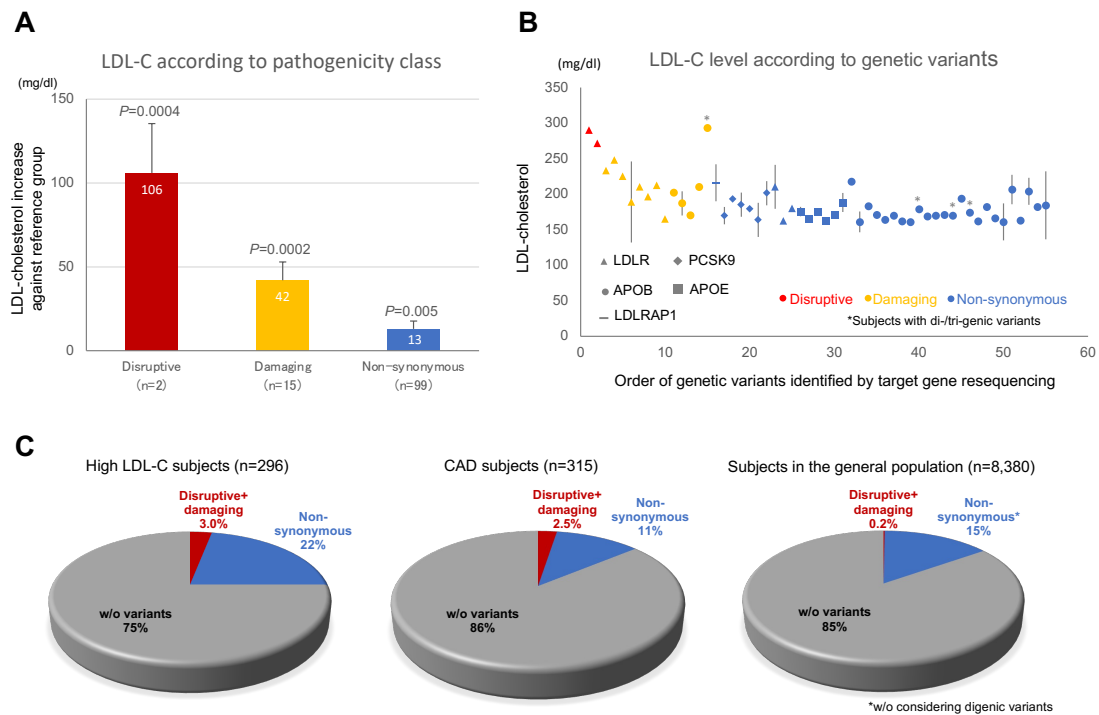


Fig. 2. Impacts of rare FH-related gene variants on LDL-C level and their prevalence.

An LDL-C increase against the reference group (BIO-CVD subjects without notable rare variants, $N = 305$) is estimated for three categories, i.e., disruptive, damaging, and non-synonymous (see Methods about the classification), of 5 FH-related gene variants (**A**), based on LDL-C level according to rare variants that were identified by target gene resequencing (**B**). **C**: The percentage of ‘disruptive + damaging’ variant carriers is shown for populations with different morbidity statuses: high LDL-C (left) and CAD (middle) subgroups and a general population (right; TMM cohort) of Japanese descent.

Fig.3

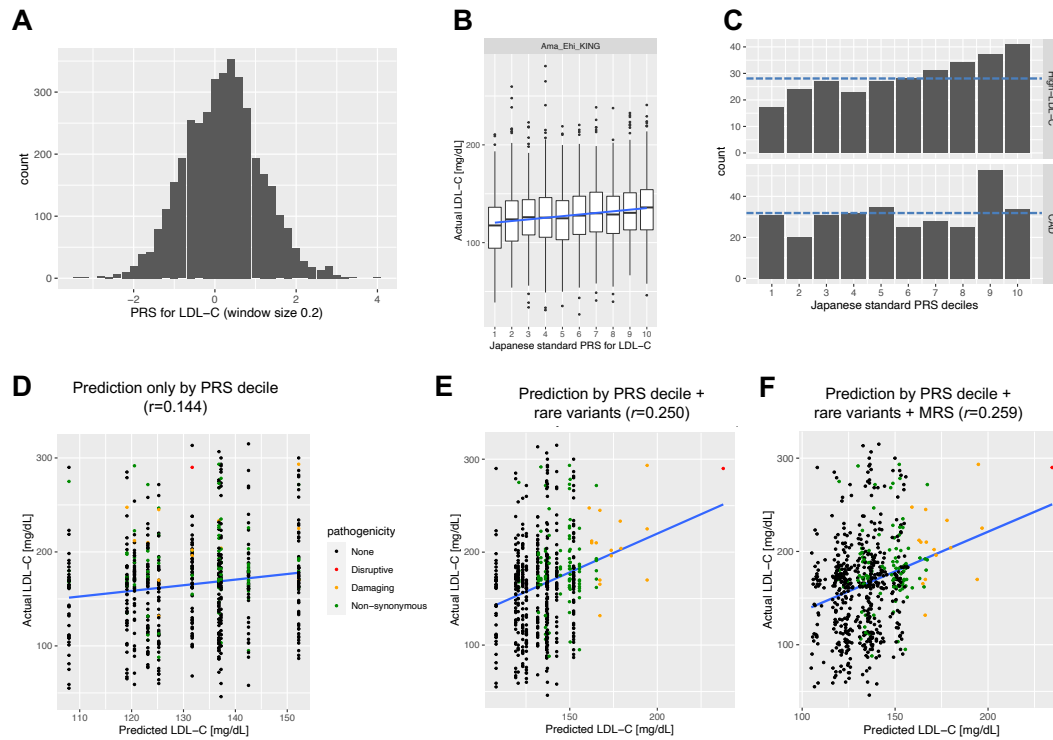


Fig. 3. Development of a conversion formula from PRS to LDL-C value for PRS-based risk prediction.

First, to define the standardized PRS_{LDL-C} decile classes, data for three population-based cohorts (Kita-Nagoya, Amagasaki, and Ehime) were combined; distribution of PRS_{LDL-C} (A) and actual LDL-C for each PRS_{LDL-C} decile class thus standardized (B). Then, data for another hospital-based study cohort (NCGM hospital cohort) were used to convert PRS to LDL-C value for the individuals newly recruited at NCGM for rare variant search besides PRS and MRS calculations (see also Fig. S2). C: PRS_{LDL-C} distribution is stratified based on standardized PRS_{LDL-C} decile classes for high LDL-C (top) and CAD (bottom) subgroups at NCGM. Correlations are shown between actual LDL-C levels and those predicted by the standardized PRS_{LDL-C} decile class alone (D), composite genetic risk (E), and composite genetic risk + MRS (F) in newly recruited individuals at NCGM (BIO-CVD + NCGM Biobank) with a complete dataset.

Fig.4

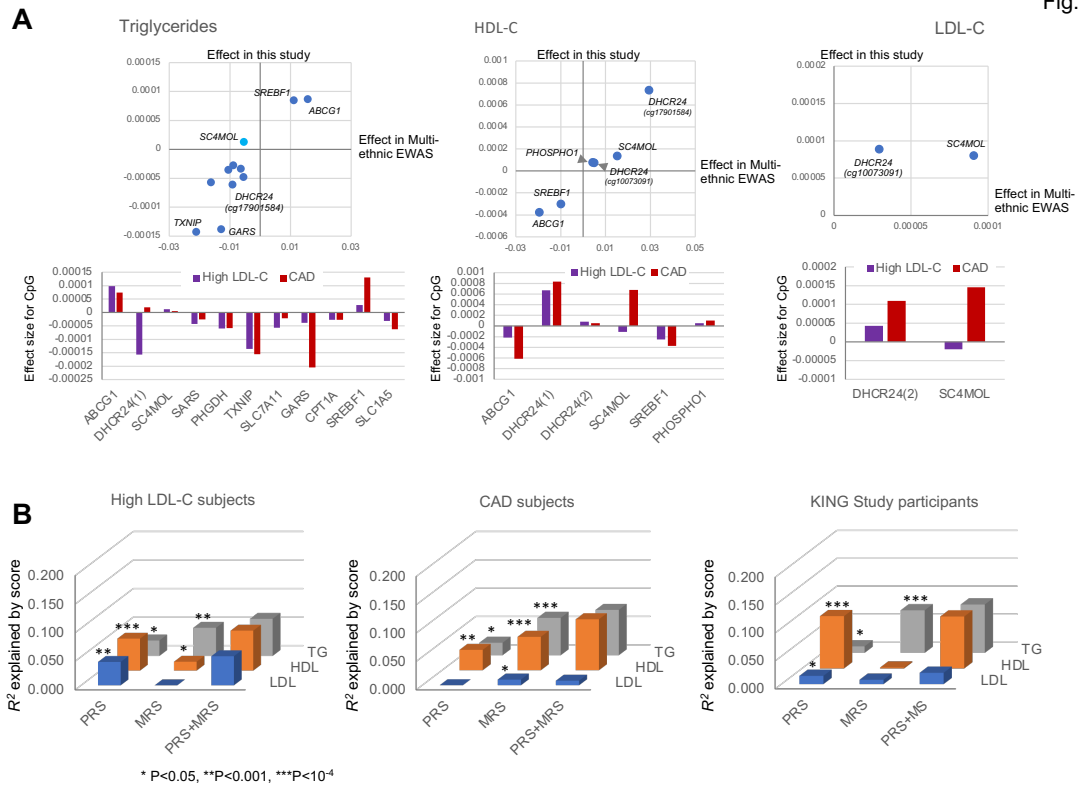


Fig. 4. CpG-trait association and prediction of lipid traits by risk scores.

A: Effect sizes for CpG-trait association are shown for triglycerides (left; at 11 CpGs), HDL-C (middle; at 6 CpGs), and LDL-C (right; at 2 CpGs), where comparisons are made between this study and previous multi-ethnic EWAS¹⁰ (top panels), and between high LDL-C and CAD subgroups (bottom panels). **B:** Impacts of two risk scores, PRS and MRS, on lipid traits are estimated by R^2 in high LDL-C (left) and CAD (middle) subgroups and a general population (right; KING study cohort) of Japanese descent separately. * $P<0.05$, ** $P<0.001$, *** $P<10^{-4}$.

Table 1. Participant characteristics

	High LDL-C (≥ 160) subjects w/o CAD	CAD subjects	non-CAD subjects
No. of individuals (F/M)	296 (151/145)	315 (51/264)	164 (54/110)
Age, yr	64.0 \pm 0.7	68.7 \pm 0.6	69.7 \pm 0.7
BMI, kg/m ²	23.6 \pm 0.2	24.5 \pm 0.2	23.6 \pm 0.3
LDL-C, mg/dl	179.4 \pm 1.3	98.6 \pm 1.7	109.5 \pm 2.1
HDL-C, mg/dl	60.4 \pm 1.1	48.7 \pm 0.7	56.7 \pm 1.1
Triglycerides, mg/dl	144.3 \pm 4.1	143.8 \pm 4.4	129.4 \pm 5.5
Smoking habit*			
Never, n (%)	142 (48%)	90 (29%)	61 (37%)
Former, n (%)	113 (38%)	153 (49%)	76 (46%)
Current, n (%)	41 (14%)	70 (22%)	27 (16%)
Complication**			
Hypertension, n (%)	153 (52%)	245 (78%)	114 (70%)
Diabetes, n (%)	52 (18%)	130 (41%)	46 (28%)
Hyper-LDL-cholesterolemia, n (%)	296 (100%)	152 (48%)	48 (29%)
Statin, n (%)	57 (19%)	221 (70%)	47 (29%)
Ezetimibe, n (%)	13 (4%)	32 (10%)	2 (1%)
CAD, n (%)***	—	315 (100%)	0 (0%)

Samples in the table were derived from either the BIO-CVD project or the NCGM Biobank and all but 122 non-CAD individuals were used for resequencing of 5 target genes related to familial hypercholesterolemia (see Methods). CAD, coronary artery disease. LDL-C, low density lipoprotein cholesterol. Of 315 CAD patients, 157 had a clear history of myocardial infarction (MI) but 158 did not: an average number of diseased vessels, 2.0 (0.1) vs 1.8 (0.1); the age of CAD onset, 63.0 yr (1.0) vs 66.9 yr (1.0); left ventricular ejection fraction in echocardiography, 50.7% (1.0) vs 63.9% (0.4) [MI vs non-MI, SE in parentheses].

*The smoking status of 2 individuals in a CAD group is unknown.

** Hypertension and diabetes are defined when individuals are prescribed anti-hypertensive and anti-diabetic drugs, respectively. Hyper-LDL-cholesterolemia is defined when basal LDL-C

≥140 mg/dl (irrespective of treatment with lipid-lowering drugs, statin, and/or ezetimibe) has been indicated by medical records.

***Subjects in a high LDL-C group do not have apparent CAD, although detailed examinations (e.g., coronary angiography) have not been performed in the majority of cases.

Table 2. A list of functional variants identified by target resequencing of FH-related genes

Chr:Pos (GRCh37, hg19)	SNP	Ref/Alt allele	Gene Name	Sequence ontology ^a	Clinical Signifi- cance ^b	N of 6, Predicted as Deleterious ^c	FATHM M MKL Coding Pred (C) ^d	LOVD classifi- cation ^e	Alt allele_freq.		χ^2 test, P-value	gnomAD PopMax ^f	LDL-C (mg/dl)	Notes
									Control (ToMMo; N = 8380)	HL+CAD (N = 610)				
Disruptive variants														
19:11215971	rs879254510	-/C	LDLR	frameshift	P/LP	—	—	P (ACGS)	N/A	0.0008	2E-04	N/A	290	D131fs
19:11227676	rs778408161	T/C	LDLR	splice_ donor	P/LP	2 of 6	Damaging	LP (ACGS)	6E-05	0.0008	0.015	7E-06	271	Disruptive (FH_Niigata)
Damaging variants														
19:11217342	rs875989907	G/A	LDLR	missense	P/LP	6 of 6	Damaging	LP	N/A	0.0008	2E-04	6E-05	233	GnomAD_exome (Asian); 3/49010
19:11221399	rs879254753	T/A	LDLR	missense	P/LP	6 of 6	Damaging	LP	6E-05	0.0008	0.015	N/A	248	Damaging (FH_Wakayama)
19:11224019	N/A	G/A	LDLR	missense	LP	6 of 6	Damaging	LP	6E-05	0.0008	0.015	N/A	225	
19:11226885	rs746959386	C/G	LDLR	missense	P/LP	5 of 6	Damaging	LP	0.0003	0.0016	0.022	1E-04	189 ± 57	
19:11213368	N/A	C/A	LDLR	missense	N/A	6 of 6	Damaging	N/A	N/A	0.0008	2E-04	N/A	210	newly identified
19:11213417	rs749038326	G/A	LDLR	missense	Conflict- ing (LP/VUS /LB*)	6 of 6	Damaging	P, LP	N/A	0.0008	2E-04	3E-04	196	GnomAD_exome (Asian); 14/49010
19:11216066	N/A	C/T	LDLR	missense	N/A	6 of 6	Damaging	P (ACGS)	N/A	0.0008	2E-04	N/A	212	c.466_484dup
19:11227613	rs201102492	G/A	LDLR	missense	Conflict- ing (P/LP/V US)	6 of 6	Damaging	LP (ACGS)	6E-05	0.0008	0.015	2E-04	165	
2:21228367	N/A	G/T	APOB	missense	N/A	4 of 6	Damaging	N/A	N/A	0.0008	2E-04	N/A	202	newly identified

2:21233912	rs759277505	G/A	APOB	missense	N/A	5 of 6	Damaging	N/A	0.00018	0.0016	0.0004	1E-05	187 ± 17	A subject possessing rs752849346 (LDLRAP1) and rs564427867 (PCSK9) concomitantly.
2:21236137	rs780170292	C/T	APOB	missense	VUS	5 of 6	Damaging	N/A	6E-05	0.0008	0.015	3E-04	170	
2:21225680	rs375471570	G/A	APOB	missense	VUS	4 of 6	Damaging	N/A	6E-05	0.0008	0.015	1E-04	210	
2:21229473	rs199689957	C/T	APOB	missense	VUS	2 of 6	Damaging	N/A	N/A	0.0008	2E-04	2E-04	293	GnomAD_exome (Asian); 0/49006. A subject possessing rs564427867 (PCSK9) concomitantly.
Gain-of-Function variants														
1:55521767	N/A	TGCCAGCGCCT/-	PCSK9	frameshift		—	—	—	N/A	0.0008	2E-04	N/A	46	NP_777596.2:p.Cys301Glyfs Ter? newly identified

^aThe highest priority ontology found among the variant–transcript interactions. The terms used are the standard feature descriptions given by the Sequence Ontology Project.

^bClinical significance is the consensus interpretation of the submissions for all conditions for this variant. Primarily based on ACMG Classifications: P, pathogenic; LP, likely pathogenic; VUS, variant of uncertain significance; LB, likely benign; B, benign.

^cVariant effect predictor tools tested are: SIFT, PolyPhen2, Mutation Taster, Mutation Assessor, FATHMM, and FATHMM-MKL.

^dIf a FATHMM-MKL coding score (PMID: 25583119) is >0.5 (or rank score >0.28317), the corresponding nsSNV is predicted as "DAMAGING"; otherwise it is predicted as "TOLERATED".

^eVariant classification and interpretation by the Association for Clinical Genomic Science (ACGS) is shown when applicable. For abbreviations, refer to the ACMG Classifications above.

^fPopMax refers to the gnomAD subpopulation with the highest allele frequency

*Despite one submission (SCV000503122.1) indicating LB, the others claim either P (6 submissions) or LP (3 submissions) for the pathogenicity of this variant.