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

Fumihiko Takeuchi, PhD\textsuperscript{1,2}; Masaya Yamamoto, MD\textsuperscript{3}; Masahiro Nakatochi, PhD\textsuperscript{4}; Kozue Takano, PhD\textsuperscript{2,5}; Atsuko Okazaki, MD\textsuperscript{5}; Sakurako Emoto, MD\textsuperscript{3}; Yasuharu Tabara, PhD\textsuperscript{6}; Tomohiro Katsuya, MD\textsuperscript{7,8}; Ken Yamamoto, MD\textsuperscript{9}; Masato Isono, PhD\textsuperscript{1}; Kotaro Mori, PhD\textsuperscript{2}; Tatsuaki Matsubara, MD\textsuperscript{10}; Sahoko Ichihara, MD\textsuperscript{11}; Mitsuhiro Yokota, MD\textsuperscript{12}; Hisao Hara, MD\textsuperscript{3}; Yukio Hiroi, MD\textsuperscript{3}; Norihiro Kato, MD\textsuperscript{1,2,5*}

\textsuperscript{1}Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine
\textsuperscript{2}Medical Genomics Center, Research Institute, National Center for Global Health and Medicine
\textsuperscript{3}Department of Cardiology, Center Hospital, National Center for Global Health and Medicine
\textsuperscript{4}Public Health Informatics Unit, Nagoya University Graduate School of Medicine
\textsuperscript{5}Department of Genomic Medicine, Center Hospital, National Center for Global Health and Medicine
\textsuperscript{6}Graduate School of Public Health, Shizuoka Graduate University of Public Health
\textsuperscript{7}Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine
\textsuperscript{8}Department of Geriatric Medicine and Nephrology, Osaka University Graduate School of Medicine
\textsuperscript{9}Department of Medical Biochemistry, Kurume University School of Medicine
\textsuperscript{10}Faculty of Human Sciences, Aichi Mizuho College
\textsuperscript{11}Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine
\textsuperscript{12}Kurume University School of Medicine

Short title: Genetic and epigenetic risk prediction of lipids

* Corresponding author

Norihiro Kato

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

Email: nokato@ri.ncgm.go.jp

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

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract (limited to 250 words)

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

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

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

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

Key Words: Genetic risk score; Plasma lipids; Coronary artery disease; DNA methylation
Dyslipidemia is a significant risk factor for atherosclerotic cardiovascular diseases, including coronary artery disease (CAD), to which genetic, environmental, and demographic factors contribute interactively. Epidemiological studies have shown that appropriate management of dyslipidemia can significantly decrease cardiovascular morbidity and mortality. Atherosclerosis treatment has often focused on reducing low-density lipoprotein cholesterol (LDL-C) among dyslipidemias, with lifestyle modifications emphasized in all patients with high LDL-C. Above all, dietary intervention is essential because high saturated fat intake causes increased concentrations of LDL-C. However, sustaining lifestyle modifications over time is generally challenging, so drug therapy is sometimes necessary to reduce the risk of atherosclerosis.

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

An individual's lipid levels are influenced by the degree of environmental exposure and genetic predisposition, although no method has been established to provide quantitative indicators of these combinations. On the one hand, it is difficult to quantitatively assess the impact of unhealthy lifestyle habits on lipid levels. Even so, exposure to some external factors, such as smoking and high plasma lipids, can cause long-lasting DNA methylation changes in blood cells. These changes can reflect an individual's environmental history, and consequently, lifestyle habits may be evaluated through epigenetic biomarkers. On the other hand, genetic information as an auxiliary diagnostic tool for dyslipidemia is still in the realm of research, with a few exceptions, such as the genetic diagnosis of FH. In this line, polygenic risk score (PRS) has recently become popular to summarize the cumulative effects of genetic loci, successfully identified by genome-wide association studies (GWAS) of lipid traits. Nonetheless, there remains considerable uncertainty in individual PRS estimation for PRS-
based risk stratification\textsuperscript{11} and substantial difficulty in unbiased assessment of the contribution of rare variants to complex traits.\textsuperscript{12}

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

Materials and Methods

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

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

In NCGM, adult patients of Japanese descent were recruited via two separate projects: BIO-CVD and NCGM Biobank, whose enrollment started in 2014 and 2012, respectively. They were categorized into three subgroups based on their plasma LDL-C levels and CAD status (see Supplementary Methods): high LDL-C without CAD, CAD, and non-CAD subgroups (as shown in Table 1). A hospital-based study called BIO-CVD assessed CAD in 475 participants (306 CAD and 169 non-CAD). Samples from the NCGM Biobank were also examined, specifically those with elevated plasma LDL-C (≥160 mg/dl) (N = 300). These samples were used for genetic and DNA methylation analyses (Fig. 1). Also, a significant number of Japanese individuals from non-targeted populations were included. We used publicly available whole genome sequencing data (N = 8.3K) for low-frequency (or rare) genetic
variants. For common single nucleotide variants (SNVs), we used GWAS array genotyping data in three population-based cohorts (including the Kita-Nagoya Genomic Epidemiology [KING] study cohort\textsuperscript{15}) and another hospital-based sample at NCGM (named NCGM hospital cohort; \(N = 351\)) (Table S1), which we referred to hereafter as reference individuals in total (\(N = 3,801\)). Furthermore, for DNA methylation, we utilized epigenome-wide association study (EWAS)-array assayed data for part (\(N = 314\)) of the KING study cohort samples previously reported.\textsuperscript{15}

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

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

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

**Gene sequencing:** Next-generation sequencing screened mutations in five FH-related genes (\(LDLR, APOB, PCSK9, LDLRAP1,\) and \(APOE)\). The sequences were aligned to a human genome reference sequence (GRCh37/hg19). SNVs and insertion/deletion (Indel) were called using the software GATK (https://gatk.broadinstitute.org/) and McCortex (https://github.com/mcveanlab/mccortex) and annotated using VarSeq (https://www.goldenhelix.com/products/VarSeq/). We then retrieved rare, putatively pathogenic variants, including SNVs that cause non-synonymous, nonsense, or splice-site substitutions and are predicted to be deleterious by ClinVar, LOVD, or in silico prediction algorithms in dbNSFP.\textsuperscript{17}
**Genotyping:** 767 samples were newly genotyped with the Infinium OmniExpress-24 BeadChip (Illumina, San Diego, CA, USA). Besides, GWAS array genotyping data for reference individuals were utilized.

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

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

**Genetic composite risk:** For rare [minor allele frequency (MAF) \(<0.01\)] variants of putative functional significance, we arbitrarily classified them into three distinct categories—(1) disruptive (frameshift or splice-donor), (2) damaging, and (3) all other non-synonymous variants—similar to the previous study with some modifications (see the Supplementary Methods). We estimated the net increase in LDL-C levels for rare variant carriers by subtracting a predicted value (based on PRS\(_{LDL-C}\)) from the actual measured value. Using linear regression models, we calculated LDL-C changes from individuals without rare FH-related gene variants and the standard error (SE) by category, including disruptive, damaging, and other non-synonymous variants.
To combine the genetic risk of rare variants and SNVs, we used the following algorithm: [an individual's LDL-C value] = [predicted LDL-C value based on PRS_{LDL-C}] + [estimated changes from non-carriers (by rare variant category)].

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

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

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

No formal sample size calculations were performed since both BIO-CVD and NCGM Biobank are...
ongoing studies and there is no generally accepted approach to estimating the sample size needed to develop a prediction model like this study.

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

**Results**

**Imputation of baseline LDL-C level**

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

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

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

We evaluated the effect sizes of rare FH-related gene variants compared to a reference group of non-carriers by category to calculate the net increase in LDL-C caused by individual rare variants (see Methods). LDL-C levels increased significantly ($P<0.05$) in all three categories compared to the reference group: plus 106, 42, and 13 mg/dl for disruptive, damaging, and non-synonymous rare variants, respectively (Fig. 2A). The order of increase was consistent with the predicted pathogenicity.
of identified rare variants (Fig. 2B) and also with results from another study in Japanese.\textsuperscript{22}

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

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

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

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

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

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

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

When the CpGs were analyzed individually, the CpG-trait association was more robust in the CAD subgroup compared to the high LDL-C subgroup for lipid traits (Fig. 4A). This was supported by the results of the MRS-trait association, which involves the cumulative effects of top-hit CpGs previously reported for each lipid trait (Fig. 4B and Table S4). Contrarily, the impact of PRSs on lipid traits tended to be smaller in the CAD subgroup than in the high LDL-C subgroup. The two risk scores
were also explanatory factors for the variance of lipid traits in the general population (as represented by the KING study cohort). At the same time, there were some differences in their relative influence compared to the high LDL-C and CAD subgroups. Overall, within each subgroup or population, the effects of PRSs and MRSs on lipid traits were additive (Fig. 4B). Although it remains a preliminary finding given the modest sample size in the current non-CAD subgroup, the strength of association between the risk scores and lipid traits appeared to be influenced by CAD status or its related factors (Table S4).

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

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

**Discussion**

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

As a first step in this study, we have successfully developed an algorithm to predict composite genetic risk for high LDL-C levels. Prioritizing the detection of rare pathogenic variants in population screening, as well as the concurrent use of PRS is a topic of active discussion for genomic conditions that are actionable and adult-onset. Thus, a study has shown a risk prediction model that combines rare pathogenic variants, PRS, and individual risk factor variables for breast cancer (i.e., dichotomous
However, this has not been previously known for lipids, a quantitative trait. The heritability of blood lipid levels, based on SNPs, is estimated to be relatively high (23.3% for LDL-C, 30.5% for triglycerides, and 32.8% for HDL-C); however, the susceptibility loci identified so far can only partially explain it. Among the lipid traits, causative gene variants with a strong genetic influence, or monogenic mutations are known principally for LDL-C and are applied in current clinical practice for the genetic diagnosis of FH. In a population-based sample of European descent, only 2% of those with severe hypercholesterolemia (LDL-C > 183 mg/dl) were reported to carry a monogenic mutation, and 23% had a high PRS (top 5th percentile) for LDL-C, indicating the importance of assessing both monogenic and polygenic models simultaneously. On the other hand, in a hospital-based setting, only about 40% of patients clinically diagnosed with FH were estimated to have an FH-causing rare variant, with the remainder likely to be polygenic in origin. Genetic effects on LDL-C levels have also been studied in relation to myocardial infarction susceptibility, with similar results to the present study for the FH-related genes LDLR and PCSK9. Accordingly, the newly developed algorithm for predicting LDL-C levels is expected to optimize preventive therapy for CAD patients.

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

Given these circumstances, we have evaluated lipid MRSs containing a list of 13-CpGs in combination with PRS or without PRS through the development of prediction models for individual trait values. We initially selected the 13 CpGs because they showed a strong association with any of
the lipid traits in the multi-ethnic EWAS, but these CpG-trait associations were reported to extend beyond lipids, considerably overlapping with other cardiometabolic phenotypes such as blood pressure, alcohol intake, and liver fat (Table S5). This leads to a hypothesis that some upstream environmental determinant (e.g., diet or exercise) is responsible for the overlap. For lipids, the effect sizes for single CpG-trait association between the present study in Japanese and previous multi-ethnic EWAS correlated well (Fig. 4A), which corroborates the reported high correlation between Europeans and non-Europeans (e.g., African Americans). Although there are some differences between subgroups in the interindividual variance explained by the MRS, the MRS predicts lipid traits in an additive manner independent of the PRS, particularly prominent among CAD subjects (Figure 4B and Table S4). This is in accord with previous findings on BMI, suggesting that the MRS represents environmental effects.

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

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

There are strengths and limitations in this study. First, we investigated the impact of genetic and DNA methylation variations on dyslipidemia in two risk groups (high LDL-C and CAD subjects), in contrast to the general population and non-CAD subjects. It is a strength that this information can be used to optimize treatment for primary (high LDL-C) and secondary (CAD) prevention of atherosclerotic cardiovascular disease by considering the relative contributions of genetics and
environmental exposure. However, given age differences, individual differences in various lifestyle habits, and the impact of drug therapy on blood lipid measurements, data collected over time (including pretreatment measurements) in larger populations should be used to improve the accuracy of dyslipidemia risk prediction. Second, although we used 13 CpGs selected from top-hit CpGs in the multi-ethnic EWAS of lipids, the method for determining a set of CpGs for the MRS is still largely undefined. To improve the precision and practicality of the prediction model, an increased number of methylation sites for MRS and standard procedures for optimal weighting and clamping of target CpG sites are needed. Third, although the interaction term was included in multivariate regression analysis, potential interactions between MRS and PRS remain to be deeply explored. Fourth, combining rare FH-related gene variants with PRS can improve the predictability of LDL-C. Still, it must be more feasible and cost-effective to achieve similar predictability for HDL-C and triglycerides in clinical practice. Nevertheless, as the cost of sequencing decreases and genetic variation annotation becomes more manageable, a similar approach for the latter two lipid traits will be justifiable. Fifth, we utilized Japanese original reference panels with individual-level data for SNPs and the phenotypic trait of interest to standardize the PRS, which requires an ancestry-matched reference. For MRS, external weights from the multi-ethnic EWAS were used to calculate it, after confirming a high correlation between ethnic groups. However, further investigation is necessary to improve the generalizability of our risk prediction index. Sixth, given the limited number of CAD patients in the general Japanese population, the current findings need to be further validated and replicated with a larger, more representative population independently.

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

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

Sources of funding:

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

Disclosures: None

Data availability:

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


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 ($N \approx 164$) 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. 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. 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. 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 EWAS\(^\text{10}\) (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

<table>
<thead>
<tr>
<th></th>
<th>High LDL-C (≥160) subjects w/o CAD</th>
<th>CAD subjects</th>
<th>non-CAD subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of individuals (F/M)</td>
<td>296 (151/145)</td>
<td>315 (51/264)</td>
<td>164 (54/110)</td>
</tr>
<tr>
<td>Age, yr</td>
<td>64.0 ± 0.7</td>
<td>68.7 ± 0.6</td>
<td>69.7 ± 0.7</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>23.6 ± 0.2</td>
<td>24.5 ± 0.2</td>
<td>23.6 ± 0.3</td>
</tr>
<tr>
<td>LDL-C, mg/dl</td>
<td>179.4 ± 1.3</td>
<td>98.6 ± 1.7</td>
<td>109.5 ± 2.1</td>
</tr>
<tr>
<td>HDL-C, mg/dl</td>
<td>60.4 ± 1.1</td>
<td>48.7 ± 0.7</td>
<td>56.7 ± 1.1</td>
</tr>
<tr>
<td>Triglycerides, mg/dl</td>
<td>144.3 ± 4.1</td>
<td>143.8 ± 4.4</td>
<td>129.4 ± 5.5</td>
</tr>
<tr>
<td>Smoking habit*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never, n (%)</td>
<td>142 (48%)</td>
<td>90 (29%)</td>
<td>61 (37%)</td>
</tr>
<tr>
<td>Former, n (%)</td>
<td>113 (38%)</td>
<td>153 (49%)</td>
<td>76 (46%)</td>
</tr>
<tr>
<td>Current, n (%)</td>
<td>41 (14%)</td>
<td>70 (22%)</td>
<td>27 (16%)</td>
</tr>
<tr>
<td>Complication**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
<td>153 (52%)</td>
<td>245 (78%)</td>
<td>114 (70%)</td>
</tr>
<tr>
<td>Diabetes, n (%)</td>
<td>52 (18%)</td>
<td>130 (41%)</td>
<td>46 (28%)</td>
</tr>
<tr>
<td>Hyper-LDL-cholesterolemia, n (%)</td>
<td>296 (100%)</td>
<td>152 (48%)</td>
<td>48 (29%)</td>
</tr>
<tr>
<td>Statin, n (%)</td>
<td>57 (19%)</td>
<td>221 (70%)</td>
<td>47 (29%)</td>
</tr>
<tr>
<td>Ezetimibe, n (%)</td>
<td>13 (4%)</td>
<td>32 (10%)</td>
<td>2 (1%)</td>
</tr>
<tr>
<td>CAD, n (%)***</td>
<td>—</td>
<td>315 (100%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Chr:Pos (GRCh37, hg19)</th>
<th>SNP</th>
<th>Ref/Alt allele</th>
<th>Gene Name</th>
<th>Sequence ontology</th>
<th>Clinical Significance</th>
<th>N of 6, Predicted as Deleterious</th>
<th>FATHMM MKL Coding Pred (C)</th>
<th>LOVD classification</th>
<th>Alt allele_freq.</th>
<th>Control (ToMMo; N = 8380)</th>
<th>HL+CAD (N = 610)</th>
<th>( \chi^2 ) test, P-value</th>
<th>gnomAD PopMax ( \times 10^6 )</th>
<th>LDL-C (mg/dl)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Disruptive variants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19:11215971</td>
<td>rs879254510</td>
<td>/C</td>
<td>LDLR</td>
<td>frameshift</td>
<td>P/LP</td>
<td>—</td>
<td>—</td>
<td>P (ACGS)</td>
<td>N/A</td>
<td>0.0008</td>
<td>2E-04</td>
<td>N/A</td>
<td>2E-04</td>
<td>290</td>
<td>D131fs</td>
</tr>
<tr>
<td>19:11227676</td>
<td>rs778408161</td>
<td>T/C</td>
<td>LDLR</td>
<td>splice_ donor</td>
<td>P/LP</td>
<td>2 of 6</td>
<td>Damaging</td>
<td>LP (ACGS)</td>
<td>6E-05</td>
<td>0.0008</td>
<td>0.015</td>
<td>7E-06</td>
<td>271</td>
<td>Disruptive (FH_Niigata)</td>
<td></td>
</tr>
<tr>
<td><strong>Damaging variants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19:11217342</td>
<td>rs875989907</td>
<td>G/A</td>
<td>LDLR</td>
<td>missense</td>
<td>P/LP</td>
<td>6 of 6</td>
<td>Damaging</td>
<td>LP</td>
<td>N/A</td>
<td>0.0008</td>
<td>2E-04</td>
<td>6E-05</td>
<td>233</td>
<td>GnomAD_exome (Asian); 3/49010</td>
<td></td>
</tr>
<tr>
<td>19:11221399</td>
<td>rs879254753</td>
<td>T/A</td>
<td>LDLR</td>
<td>missense</td>
<td>P/LP</td>
<td>6 of 6</td>
<td>Damaging</td>
<td>LP</td>
<td>6E-05</td>
<td>0.0008</td>
<td>0.015</td>
<td>N/A</td>
<td>248</td>
<td>Damaging (FH_Wakayama)</td>
<td></td>
</tr>
<tr>
<td>19:11224019</td>
<td>N/A</td>
<td>G/A</td>
<td>LDLR</td>
<td>missense</td>
<td>P/LP</td>
<td>6 of 6</td>
<td>Damaging</td>
<td>LP</td>
<td>6E-05</td>
<td>0.0008</td>
<td>0.015</td>
<td>N/A</td>
<td>225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19:11226885</td>
<td>rs746959386</td>
<td>C/G</td>
<td>LDLR</td>
<td>missense</td>
<td>P/LP</td>
<td>5 of 6</td>
<td>Damaging</td>
<td>LP</td>
<td>0.0003</td>
<td>0.016</td>
<td>0.022</td>
<td>1E-04</td>
<td>189 ± 57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19:11213368</td>
<td>N/A</td>
<td>C/A</td>
<td>LDLR</td>
<td>missense</td>
<td>N/A</td>
<td>6 of 6</td>
<td>Damaging</td>
<td>N/A</td>
<td>N/A</td>
<td>0.0008</td>
<td>2E-04</td>
<td>N/A</td>
<td>210</td>
<td>newly identified</td>
<td></td>
</tr>
<tr>
<td>19:11213417</td>
<td>rs749038326</td>
<td>G/A</td>
<td>LDLR</td>
<td>missense</td>
<td>Conflict ing (LP/VUS/LB*)</td>
<td>6 of 6</td>
<td>Damaging</td>
<td>P, LP</td>
<td>N/A</td>
<td>0.0008</td>
<td>2E-04</td>
<td>3E-04</td>
<td>196</td>
<td>GnomAD_exome (Asian); 14/49010</td>
<td></td>
</tr>
<tr>
<td>19:11216066</td>
<td>N/A</td>
<td>C/T</td>
<td>LDLR</td>
<td>missense</td>
<td>N/A</td>
<td>6 of 6</td>
<td>Damaging</td>
<td>P (ACGS)</td>
<td>N/A</td>
<td>0.0008</td>
<td>2E-04</td>
<td>N/A</td>
<td>212</td>
<td>c.466_484dup</td>
<td></td>
</tr>
<tr>
<td>19:11227613</td>
<td>rs201102492</td>
<td>G/A</td>
<td>LDLR</td>
<td>missense</td>
<td>Conflict ing (P/LP/V US)</td>
<td>6 of 6</td>
<td>Damaging</td>
<td>LP (ACGS)</td>
<td>6E-05</td>
<td>0.0008</td>
<td>0.015</td>
<td>2E-04</td>
<td>165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:21228367</td>
<td>N/A</td>
<td>G/T</td>
<td>APOB</td>
<td>missense</td>
<td>N/A</td>
<td>4 of 6</td>
<td>Damaging</td>
<td>N/A</td>
<td>N/A</td>
<td>0.0008</td>
<td>2E-04</td>
<td>N/A</td>
<td>202</td>
<td>newly identified</td>
<td></td>
</tr>
<tr>
<td>Chromosome</td>
<td>rsID</td>
<td>SNP Type</td>
<td>Gene</td>
<td>Coding Effect</td>
<td>Annotation</td>
<td>Prioritization</td>
<td>p-value</td>
<td>q-value</td>
<td>Experimental p-value</td>
<td>Clinical Significance</td>
<td>PopMax</td>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
<td>--------------</td>
<td>------------</td>
<td>---------------</td>
<td>---------</td>
<td>---------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>--------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:21233912</td>
<td>rs759277505</td>
<td>G/A</td>
<td>APOB</td>
<td>missense</td>
<td>N/A</td>
<td>5 of 6</td>
<td>Damaging</td>
<td>N/A</td>
<td>0.0018</td>
<td>0.0016</td>
<td>0.0004</td>
<td>1E-05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:21236137</td>
<td>rs780170292</td>
<td>C/T</td>
<td>APOB</td>
<td>missense</td>
<td>VUS</td>
<td>5 of 6</td>
<td>Damaging</td>
<td>N/A</td>
<td>6E-05</td>
<td>0.0008</td>
<td>0.015</td>
<td>1E-04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:21225680</td>
<td>rs375471570</td>
<td>G/A</td>
<td>APOB</td>
<td>missense</td>
<td>VUS</td>
<td>4 of 6</td>
<td>Damaging</td>
<td>N/A</td>
<td>6E-05</td>
<td>0.0008</td>
<td>0.015</td>
<td>1E-04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:21229473</td>
<td>rs199689957</td>
<td>C/T</td>
<td>APOB</td>
<td>missense</td>
<td>VUS</td>
<td>2 of 6</td>
<td>Damaging</td>
<td>N/A</td>
<td>2E-04</td>
<td>2E-04</td>
<td>293</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Gain-of-Function variants**

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>rsID</th>
<th>SNP Type</th>
<th>Gene</th>
<th>Coding Effect</th>
<th>Annotation</th>
<th>Prioritization</th>
<th>p-value</th>
<th>q-value</th>
<th>Experimental p-value</th>
<th>Clinical Significance</th>
<th>PopMax</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:55521767</td>
<td>N/A</td>
<td>TGCAAGGAGCT/</td>
<td>PCSK9</td>
<td>frameshift</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>46</td>
<td></td>
</tr>
</tbody>
</table>

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

**Clinical 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.

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

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

**PopMax 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.*