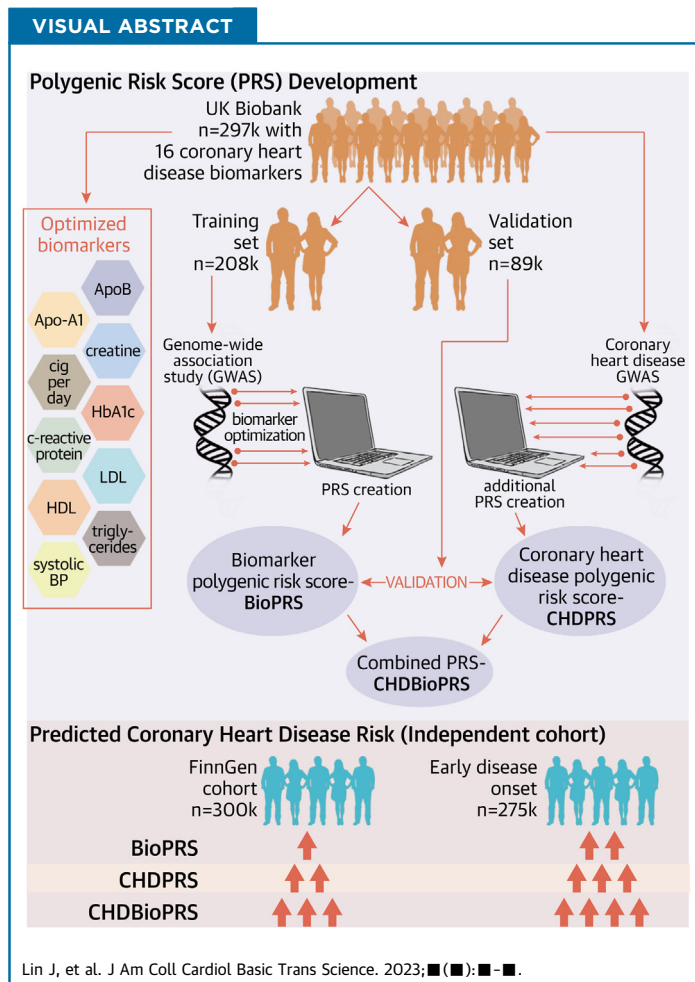


PRECLINICAL RESEARCH

Integration of Biomarker Polygenic Risk Score Improves Prediction of Coronary Heart Disease

Jake Lin, PhD,^{a,b} Nina Mars, PhD,^a Yu Fu, MSc,^a Pietari Ripatti, PhD,^a Tuomo Kiiskinen, MD,^a FinnGen study, Taru Tukiainen, PhD,^a Samuli Ripatti, PhD,^{a,c,d,e} Matti Pirinen, PhD^{a,c,f}



HIGHLIGHTS

- The novel BioPRS, constructed from combining statistically relevant CHD biomarkers, was clearly predictive of CHD in both the UK Biobank and FinnGen.
- CHDBioPRS, combining BioPRS with a standard CHD PRS, improved the prediction of the CHD PRS, with the largest effect size observed among the early onset cases.
- We observed similar HRs of CHDBioPRS for men and women.

From the ^aInstitute for Molecular Medicine Finland, Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland; ^bHealth Sciences, Faculty of Social Sciences, Tampere University, Tampere, Finland; ^cDepartment of Public Health, University of Helsinki, Helsinki, Finland; ^dBroad Institute of Massachusetts Institute of Technology, Harvard University, Cambridge, Massachusetts, USA; ^eMassachusetts General Hospital, Cambridge, Massachusetts, USA; and the ^fDepartment of Mathematics and Statistics, University of Helsinki, Helsinki, Finland.

ABBREVIATIONS AND ACRONYMS

Apo = apolipoprotein

BMI = body mass index

CHD = coronary heart disease

CPD = cigarettes per day

CRP = C-reactive protein

GWAS = genome-wide
association study

HbA_{1c} = glycated hemoglobin

HDL = high-density-lipoprotein

LD = linkage disequilibrium

LDL = low-density-lipoprotein

LRT = likelihood ratio test

MI = myocardial infarction

NRI = net reclassification
improvement

PRS = polygenic risk score

SBP = systolic blood pressure

SNP = single nucleotide
polymorphism

TRIG = triglycerides

UKB = UK Biobank

SUMMARY

There are several established biomarkers for coronary heart disease (CHD), including blood pressure, cholesterol, and lipoproteins. It is of high interest to determine how a combined polygenic risk score (PRS) of CHD-associated biomarkers (BioPRS) can further improve genetic prediction of CHD. We developed CHDBioPRS, combining BioPRS with PRS of CHD in the UK Biobank and tested it on FinnGen. We found that BioPRS was clearly predictive of CHD and that CHDBioPRS improved the standard CHD PRS. The largest effect was observed with early onset cases in FinnGen, with HRs above 2 per standard deviation of CHDBioPRS. (J Am Coll Cardiol Basic Trans Science 2023; ■: ■-■) © 2023 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Coronary heart disease (CHD), a complex disease caused by a gradual build-up of fatty deposits in the arteries, is a major cause of death worldwide. In addition to family history, age, sex, smoking history, and levels of blood pressure, inflammation and lipoproteins are established biomarkers for CHD.¹⁻⁴ These risk factors are also used in clinical risk calculators to evaluate preventive therapies and strategies. Although clinical risk scores enable

identification of some individuals at high risk,⁵⁻⁷ a large proportion of CHD cases are not detected by these scores, and the utility of clinical scores is limited for young adults^{3,8-10} and women.^{11,12}

Genome-wide association studies (GWAS) involving large human genetic data sets have identified more than 100 loci statistically associated with CHD, mostly within populations of European descent.¹³⁻¹⁸ These genetic discoveries together with the introduction of sophisticated statistical tools that incorporate linkage disequilibrium information, have greatly advanced risk prediction.¹⁹⁻²¹ Particularly, several studies have shown that inclusion of polygenic risk scores (PRS) improve CHD prediction and identification of high-risk groups.^{13,22-24} Because PRS are based on germline DNA, risk profiling can be conducted in early life when the individuals with the highest values of PRS are likely to benefit from an early adoption of preventive strategies.

The landmark study of PRS for common diseases conducted by Khera et al²³ showed that a sizable portion of the population carry a polygenic CHD risk equivalent to known monogenic mutations conferring severalfold increased risk. This PRS,

comprising more than 6 million single nucleotide polymorphisms (SNPs), was generated with the use of LDpred¹⁹ and has proven to be effective in validation sets across multiple populations while also performing favorably compared with other PRS²⁵ composed of smaller numbers of variants.

Because PRS have proven to be successful for CHD prediction, it remains of high interest to systematically determine how a combined polygenic biomarker score (BioPRS) constructed with biomarkers associated with CHD can improve on the established CHDPRS. Recently, multi-PRS models, using 35 PRS from blood and urine biomarkers, have been shown to improve genetic risk prediction of common diseases such as type 2 diabetes and gout.² While other existing studies have focused on combining several GWAS of CHD,^{13,26,27} our focus is on the combination of effects of known CHD-associated biomarkers into a single PRS and its integration with CHDPRS. Furthermore, because the current risk calculators do not work equally well for women as for men, it is of high importance to quantify the contribution of BioPRS within each sex.²⁸ Another important goal is to predict a subgroup of CHD cases with an early onset of the disease.²⁹

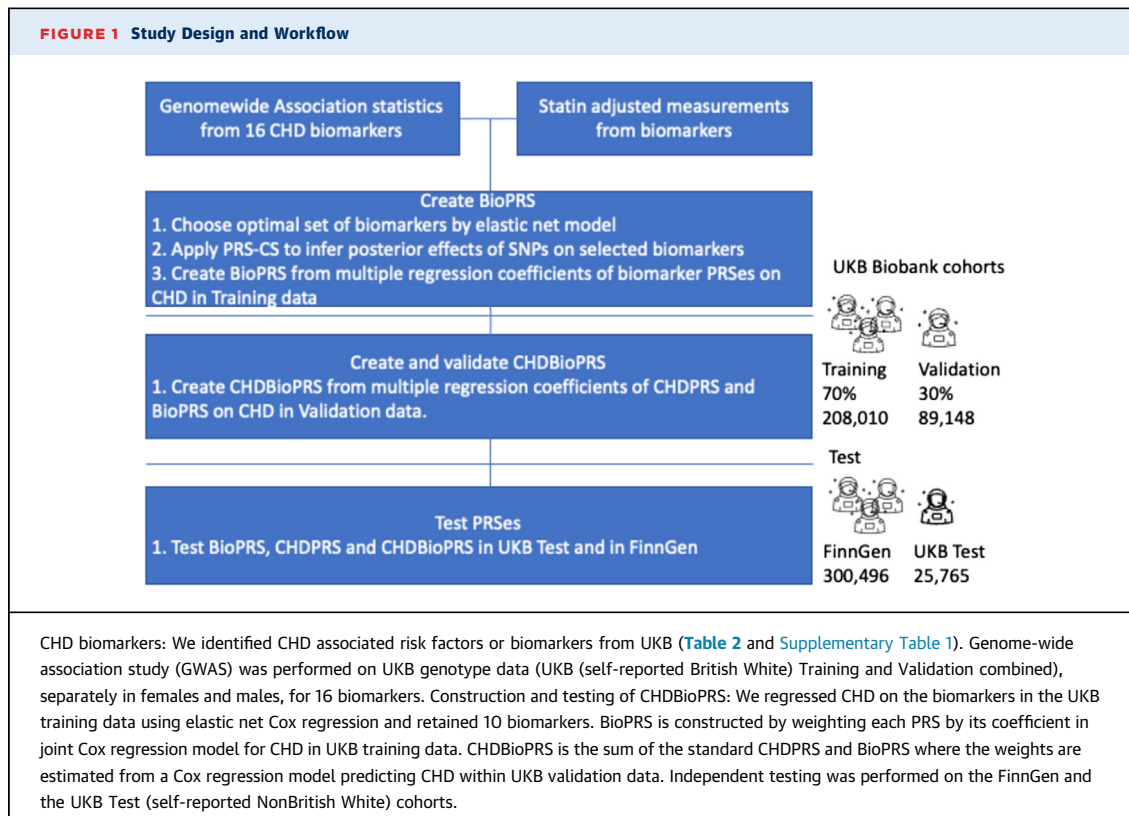
METHODS

The workflow of the study is presented in [Figure 1](#).

UK BIOBANK DATA. The design of UK Biobank (UKB) and the background of its participants have been reported previously.^{22,30} We restricted our analyses to samples that were of self-reported European ancestry to avoid potential spurious associations driven by allele frequency differences when including

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the [Author Center](#).

Manuscript received March 9, 2023; revised manuscript received July 6, 2023, accepted July 10, 2023.



individuals from different ancestry backgrounds in our GWAS. We made use of 4 sets of UKB samples. First, our GWAS set contained 343,695 samples who were unrelated (pairwise kinship coefficients reported by UKB <0.044) and who self-reported “White British” as their ethnicity. Out of the GWAS set, 297,158 individuals had data on all 16 biomarkers (listed subsequently) and we split these into our UKB training (70%, 208,010 individuals) and UKB validation (30%, 89,148 individuals) sets. The split was done randomly by maintaining a constant CHD case ratio and sex ratio within each set. Finally, we also collected a UKB test set of 25,765 unrelated individuals with self-reported ethnicity as “White non-British” and biomarker measurements available. The UKB test set did not overlap with the UKB GWAS set, UKB training set, or UKB validation set.

UKB CHD DEFINITION. The UKB CHD endpoint was defined as fatal or nonfatal myocardial infarction (MI), or coronary revascularization (percutaneous transluminal coronary angioplasty [PTCA] or coronary artery bypass grafting [CABG]). In detail, and consistently with previous studies,^{23,31} CHD cases were defined as having a heart attack diagnosed by a doctor or self-report of noncancer illness or operation including PTCA and CABG. In addition, coronary revascularization was assigned based on

OPCS-4 coded procedure for CABG (K40.1-4, K41.1-4, and K45.1-5) or PTCA (K49.1-2, K49.8-9, K50.2, K75.1-4, and K75.8-9). MI cases from summary hospital episode statistics (ICD-10 and ICD-9 codes) were defined with ICD-10 I21, I22, I23, I24.1, and I25.2 and ICD-9 410, 411, and 412. We further merged the outcomes from the UKB MI algorithmically defined outcomes and dates. The UKB field codes of the CHD endpoint are listed in Table 1. We defined prevalent CHD cases as those who had CHD already when their blood sample was taken; other CHD cases were considered incident CHD cases. The age at onset in prevalent cases was determined by hospital episode data and self-reported age of onset, or for incident CHD cases by hospital or death records demonstrating disease onset after UKB enrollment.

UKB BIOMARKERS AND GWAS. We identified 16 variables from UKB that have been previously reported to be associated with CHD (the UKB codes are listed in Table 2): high-density-lipoprotein (HDL) cholesterol, low-density-lipoprotein (LDL) direct, triglycerides (TRIG), apolipoprotein (Apo) A1, ApoB, diastolic blood pressure, systolic blood pressure (SBP), glycated hemoglobin (HbA_{1c}), glucose, C-reactive protein (CRP), creatinine, lipoprotein(a), doctor-diagnosed diabetes, body mass index (BMI) and

TABLE 1 Sample Characteristics in UKB and FinnGen

	UKB	UKB Female	UKB Male	FinnGen	FinnGen Female	FinnGen Male
Sample size	297,158	158,382	138,776	300,496	169,508	130,988
Enrollment age, y	56.9 ± 7.99	56.7 ± 7.90	57.12 ± 8.09	53.88 ± 16.63	52.35 ± 16.96	55.86 ± 15.74
Incident CHD cases	5,011 (1.69)	1,353 (0.90)	3,658 (2.63)	13,623 (4.53)	3,652 (2.15)	9,971 (7.61)
Age at onset for incident cases, y ^a	65.13 ± 6.90	66.24 ± 6.95	64.92 ± 7.14	70.29 ± 10.80	72.56 ± 12.16	69.46 ± 10.13
Follow-up, y	10.01 ± 0.90	10.00 ± 0.89	10.04 ± 0.90	10.43 ± 8.96	9.22 ± 8.72	10.89 ± 9.01
All CHD cases ^b	16,803 (5.65)	3,762 (2.38)	12,896 (9.30)	30729 (10.22)	8,486 (5.00)	22,243 (16.98)
Age at onset, y	59.93 ± 9.43	61.41 ± 9.83	59.50 ± 9.26	65.37 ± 11.82	67.97 ± 12.50	64.37 ± 11.39

Values are mean ± SD or n (%). ^aFor incident coronary heart disease (CHD) cases, age at onset is after the blood sample was taken. ^bCHD case endpoints were defined based on the following UKB (UK Biobank) fields (codes): myocardial infarction (MI) hospitalization: 41202 (ICD-10)/41203 (ICD-9); heart attack doctor diagnosed: 6150; percutaneous transluminal coronary angioplasty operation: 20004 (K49.1-2, K49.8-9, K50.2, K75.1-4, and K75.8-9); coronary revascularization: 41272 (K40.1-4, K41.1-4, and K45.1-5); and MI algorithm: 42000/42001 (date of onset).

cigarettes per day (CPD). In addition, total cholesterol was calculated from the Friedewald formula³² as HDL + LDL + TRIG/2.2 in units of mmol/L. For simplicity, we use the term “biomarker” to refer to all of these variables even though some of them (namely, blood pressure, diabetes, BMI, and CPD) are not typical biomarkers measured from a blood or tissue sample.

Statin usage was identified from treatment medication (13 drugs: 1141146234, atorvastatin; 1141192414, crestor 10 mg tablet; 1140910632, eptastatin; 1140888594, fluvastatin; 1140864592, lescol 20 mg capsule; 1141146138, lipitor 10 mg tablet; 1140861970, lipostat 10 mg tablet; 1140888648, pravastatin; 1141192410, rosuvastatin; 1141188146, simvador 10 mg tablet; 1140861958, simvastatin; 1140881748, zocor 10 mg tablet; 1141200040, zocor heart-pro 10 mg tablet). Adjustment for statin usage² was done for the

following biomarkers by dividing the biomarker value with the given coefficient: HDL 1.053, LDL 0.684, TRIG 0.874, ApoA1 1.07, ApoB 0.722, lipoprotein(a) 1.102, glucose 1.029, CRP 1.230, creatinine 1.058, and HbA_{1c} 1.042. We also adjusted the subjects taking blood pressure reduction drugs by adding +15 to SBP and +10 to diastolic blood pressure. UKB statin and blood pressure drug codes are listed in [Table 2](#), along with the statin and blood pressure-adjusted values of these biomarkers for the UKB training set.

GENOME-WIDE ASSOCIATION STUDY. Genome-wide association study (GWAS), using BOLT-LMM v2.3.2,³³ was performed on UKB genotype data (release 3), separately in 183,130 women and 157,821 men, for the 16 biomarkers. We first regressed out sex, age, age-squared, and the top 10 principal components of genetic structure from the biomarkers and then

TABLE 2 UKB Biomarkers Mean Levels After Adjustment for Statins

Biomarker	UKB Code	Control Cases (Female)	Incident CHD Cases (Female)	Statin Adjustment ^a
Apolipoprotein A1, ^b g/L	30630	1.52 ± 1.61	1.37 ± 1.51	/1.065364
Apolipoprotein B, ^b g/L	30640	1.08 ± 1.07	1.14 ± 1.15	/0.721928
Body mass index, kg/m ²	21001	27.31 ± 26.99	28.82 ± 28.52	
Cigarettes per day ^b	3456	1.05 ± 0.92	1.89 ± 2.19	
Creatinine ^b , μmol/L	30700	71.60 ± 63.98	75.38 ± 62.91	/1.0580718
C-Reactive protein, ^b mg/L	30710	2.48 ± 2.61	2.89 ± 3.54	/1.2300281
Diabetes	2443	4.1% ± 3.1%	13.1% ± 10.3%	
Diastolic BP, mm Hg	4079	84.12 ± 82.24	87.09 ± 84.97	+10 (BP meds)
Glucose, mmol/L	30740	5.07 ± 5.03	5.37 ± 5.33	/1.028824
HbA _{1c} , ^b mmol/mol	30750	35.54 ± 35.41	38.0 ± 37.53	/1.0418022
HDL, ^b mmol/L	30760	1.46 ± 1.60	1.22 ± 1.43	/1.053
Lipoprotein A, nmol/L		34.25 ± 35.17	38.59 ± 37.81	/1.101954
LDL, ^b mmol/L	30780	3.78 ± 3.80	3.90 ± 4.04	/0.684
Systolic BP, ^b mm Hg		140.8 ± 137.85	149.65 ± 148.04	+15 (BP meds)
Total cholesterol, mmol/L	NA ^c	6.06 ± 6.16	6.26 ± 6.56	
Triglycerides, ^b mmol/L	30870	1.78 ± 1.58	2.18 ± 1.96	/0.874

^aStatin adjustment for statin users (16.2%) is done either by division (/) or addition (+) by the value given in the last column. Statin and blood pressure medicinal use was identified with the use of fields 20003, 6153, and 6177. ^bSelected in optimal regularized model. ^cTotal cholesterol calculated from Friedewald formula of HDL + LDL + TRIG/2.2 in units of mmol/L.

BP = blood pressure; HDL = high-density lipoprotein; LDL = low-density lipoprotein; other abbreviations as in [Table 1](#).

applied rank-based inverse-normal transformation to the residuals.

FinnGen. The design of the Finnish FinnGen³⁴ project and participant backgrounds are presented in [Table 1](#). The FinnGen test cohort contained 321,302 FinnGen data freeze 7 participants. The CHD case definition in FinnGen (I9_CHD) is consistent with our UKB CHD definition except that the FinnGen definition also includes samples with angina only (I20.0) as cases. Consequently, we removed the 2,989 angina-only cases from FinnGen, in addition to removal of 6,109 participants younger than 16 years old at enrollment.

BIOMARKER MODEL. Using the 16 CHD associated biomarkers for the UKB training data as predictors and incident CHD as outcome (and excluding prevalent CHD cases), we used penalized Cox proportional hazard models (glmnet R package) using the elastic net penalty ($\alpha = 0.50$) with 20-fold cross-validation.^{35,36} The optimal model identified 10 biomarkers, which were selected for BioPRS construction ([Table 2](#)). As shown in [Supplemental Figure 1](#), $\alpha = 0.25$ (closer to ridge regression) produced the same optimal set as $\alpha = 0.50$, and $\alpha = 0.75$ (closer to lasso regression) further excluded ApoA1.

BIOMARKER PRS. PRS with continuous shrinkage (CS)²¹ was run on each of the GWAS summary results of the selected 10 biomarkers. To account for linkage disequilibrium (LD), we used the 1000 Genomes^{13,37} project's phase 3 European reference panel, which resulted in LD-adjusted posterior effect sizes for 1,139,910 SNPs. The PRS for each biomarker was then computed as a sum over SNPs of the products of the individual's genotype and the posterior effect size for the SNP with the use of PLINK2.0.³⁸

WEIGHTS OF BIOMARKER PRS ON CHD. The PRS of the 10 selected biomarkers were used as predictors for incident CHD in a Cox proportional hazards model in the UKB training data without the prevalent CHD cases. According to this model, the hazard rate at age t depends on the predictors as follows:

$$h(t) = h_0(t) \cdot \exp \left(b_{hdIPRS} X_{hdIPRS} + b_{ldIPRS} X_{ldIPRS} + b_{tgPRS} X_{tgPRS} + b_{apoaPRS} X_{apoaPRS} + b_{apobPRS} X_{apobPRS} + b_{cpdPRS} X_{cpdPRS} + b_{creaPRS} X_{creaPRS} + b_{cpdPRS} X_{cpdPRS} + b_{hba1cPRS} X_{hba1cPRS} + b_{sbpPRS} X_{sbpPRS} + b_z^T X_z \right) \quad (\text{Equation 1})$$

Where $h_0(t)$ is the baseline hazard rate, z denotes the vector of covariate values (sex and the first 10 principal components of population structure), and each

biomarker has coefficient $b_{\text{biomarkerPRS}}$ that corresponds to a change in the logarithm of the hazard rate per one standard deviation of the biomarker PRS value.

We used the “coef” function from the “survival” library³⁹ of R software to estimate the $b_{\text{biomarkerPRS}}$ coefficients as previously recommended.^{40,41}

BioPRS. We combined the biomarker PRS into a score named BioPRS by standardizing (mean of 0 and SD of 1) the sum of the 10 biomarker PRS after multiplying each PRS by the beta-coefficient (b_i) of the corresponding biomarker from formula (equation 1):

$$\text{BioPRS} = \text{standardize} \left(\sum_{i=1}^{10} b_i \cdot \text{PRS}_i \right) \quad (\text{Equation 2})$$

CHDPRS. We generated a PRS for CHD (named CHDPRS) by applying PRS-CS²¹ to CHD GWAS reported by Nikpay et al¹³ using the European panel from the 1000 Genomes Project³⁷ for LD reference. Our CHDPRS contained 1,087,715 SNPs. In addition, we compared our CHDPRS with “Khera PRS,” which is the PRS for CHD generated with the use of LDpred by Khera et al²³ based on the same GWAS¹³ that we used to generate our CHDPRS.

CHDBioPRS. CHDBioPRS was constructed from integration of BioPRS and CHDPRS. Weights of the 2 PRS (CHDPRS and mBioPRS) were estimated in the UKB validation set with the use of a Cox regression model with CHD as outcome:

$$h(t) = h_0(t) \cdot \exp (C_{\text{CHDPRS}} X_{\text{CHDPRS}} + C_{\text{BioPRS}} X_{\text{BioPRS}}) \quad (\text{Equation 3})$$

CHDBioPRS is the standardized sum of the CHDPRS and BioPRS multiplied by their weights from formula (equation 3):

$$\text{CHDBioPRS} = \text{standardize} (C_{\text{CHDPRS}} \text{CHDPRS}_i + C_{\text{BioPRS}} \text{BioPRS}_i) \quad (\text{Equation 4})$$

In addition to the derivation above, a similar procedure was done also for men and women separately (biomarker selection using glmnet, biomarker weights in BioPRS using Cox regression, and combination of CHDPRS and BioPRS using another Cox regression).

SCORE2. SCORE2⁴² is a prediction model for 10-year cardiovascular disease risk that uses information on age, total cholesterol, HDL, SBP, diabetes, and smoking. We calculated SCORE2 in our UKB data sets to give a comparison point for our PRS. We note that performance of SCORE2 may be overly optimistic in UKB, because the UKB data were used

TABLE 3 HRs (With 95% CIs) From Cox Regression Model of 3 Different PRS on Incident CHD

Cohort	BioPRS	CHDPRS	CHDBioPRS
UKB training	1.45 (1.42-1.48)	1.64 (1.61-1.67)	1.76 (1.72-1.79)
UKB validation	1.42 (1.38-1.46)	1.62 (1.57-1.67)	1.73 (1.68-1.78)
UKB test	1.45 (1.36-1.55)	1.78 (1.67-1.91)	1.88 (1.76-2.01)
UKB test women	1.53 (1.34-1.75)	1.72 (1.50-1.98)	1.86 (1.62-2.14)
UKB test men	1.42 (1.32-1.53)	1.72 (1.60-1.85)	1.84 (1.71-1.98)
UKB test early onset	1.60 (1.43-1.78)	1.90 (1.69-2.13)	2.07 (1.85-2.32)
FinnGen	1.27 (1.26-1.29)	1.57 (1.55-1.60)	1.60 (1.58-1.62)
FinnGen women	1.27 (1.24-1.30)	1.53 (1.50-1.57)	1.56 (1.53-1.60)
FinnGen men	1.27 (1.25-1.29)	1.58 (1.55-1.60)	1.61 (1.59-1.63)
FinnGen early onset	1.51 (1.47-1.55)	2.01 (1.95-2.07)	2.10 (2.04-2.16)

BioPRS is combination of PRS of 10 CHD-related biomarkers, CHDPRS is a standard PRS for CHD, and CHDBioPRS combines BioPRS and CHDPRS. Early onset was defined as CHD before 55 years of age. C-index, AUC, and P values are in Supplemental Tables 2 to 7, likelihood ratio statistics are in Supplemental Table 18, and SCORE2-related statistics are in Supplemental Table 19.

PRS = polygenic risk score; other abbreviations as in Table 1.

in the derivation of SCORE2. Because SCORE2 requires laboratory measurements, it cannot be applied in FinnGen, where those lab measurements are not available. We constructed combined predictors SCORE2 + CHDPRS and SCORE2 + CHDBioPRS by an approach similar to that described by equations (3) and (4).

EARLY ONSET. We identified all individuals with early CHD onset (<55 years of age), and to further account for sex differences, we defined early CHD onset for women as <60 years of age and early CHD onset for men as <50 years of age.⁴³

STATISTICAL ANALYSIS. All analyses were done with the use of R software.^{36,39,44-46} All scores were standardized to have a mean of 0 and a variance of 1. Cox proportional hazard models were used to estimate time until CHD, with results presented as the HR with 95% CI and compared with the use of the likelihood ratio test statistic (LRT). In Cox regression, age was used as the time scale. When evaluating PRS, the CHD outcome variable included both incident and prevalent cases. In UKB, we excluded prevalent cases from the analysis when we selected relevant biomarkers and when we estimated the biomarker weights for generating BioPRS. In survival analyses of early onset cases, the cases with late onset were excluded and the control cases were censored at the upper limit of the early onset age.

C-Index,⁴⁷ a metric for prediction concordance, and AUC with 95% CI were used to assess model discrimination. We also computed how much each biomarker PRS explains of the variance of the corresponding biomarker by using the adjusted R^2 measure

for the linear model where the biomarker PRS was the only predictor in the model. Net reclassification improvement (NRI) between different prediction models was obtained to determine how well the new model reclassifies patients compared with the previous model.

RESULTS

For our UKB GWAS set, we generated statin-adjusted GWAS results for 16 CHD-associated biomarkers. After accounting for withdrawals and missing biomarker measurements, we identified a total of 297,158 participants (53.3% female) comprising 16,658 CHD cases (22.6% female). We further split these data into training (70%) and validation (30%) sets while maintaining similar case and sex proportions. Our UKB test data set of 25,765 (56.4% female) unrelated individuals of “non-British White” self-identified ethnicity included 949 CHD cases (23.2% female). Our FinnGen test data contained 300,496 Finnish individuals (56.4% women; 10.4% CHD cases of which 27.6% were women). A combined PRS of CHD biomarkers (BioPRS) was derived across PRS of 10 biomarkers (Supplemental Table 1) (HDL, LDL, TRIG, total cholesterol, systolic blood pressure, CPD, HbA_{1c}, CRP, creatinine, ApoB, ApoA1) by combining the individual biomarker PRS generated with PRS-CS on 1,106,191 SNPs. We validated in UKB test data that this BioPRS (C-index: 0.781; SE: 0.007) performed at least as well as a corresponding PRS constructed from all 16 biomarkers (C-index: 0.778; SE: 0.007). BioPRS was further integrated with the CHDPRS constructed from 6,630,150 variants from a GWAS¹³ involving 184,305 participants of European ancestry to yield CHDBioPRS. We also compared our CHDPRS to the Khera PRS²³ and found that our CHDPRS gave a similar result in UKB training data (HR: 1.64 [95% CI: 1.61-1.67]; compared with HR: 1.62 (95% CI: 1.59-1.64) (Supplemental Table 2). CHDBioPRS is approximately normally distributed and, on average, higher in CHD cases relative to the control cases (mean difference: 0.518 SD; 95% CI: 0.490-0.546) (Supplemental Figure 2).

Table 3 presents the results from UKB data sets and FinnGen. In UKB, results from the training set are similar to results from the validation and test sets, which suggests that the model is not overfitting in the training data. In all 3 UKB data sets, we observed the same pattern, where BioPRS itself is clearly predictive of CHD (HR estimates per SD vary from 1.42 to 1.45), CHDPRS on its own is more predictive than BioPRS (HR estimates vary from 1.62 to 1.78), and CHDBioPRS is the most predictive (HRs vary from 1.73 to 1.88). For

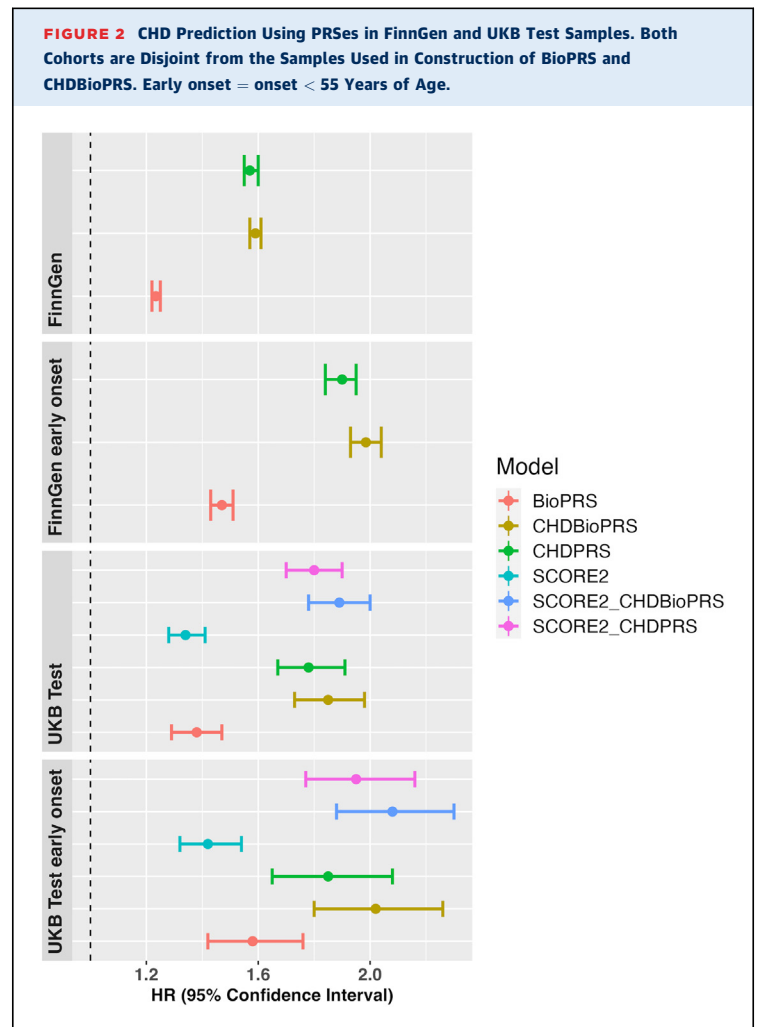
C-index, z-scores, and AUC metrics, see [Supplemental Tables 2 to 4](#). When the scores are applied in FinnGen ([Supplemental Table 5](#)), the pattern was repeated, where BioPRS is clearly predictive and improves the prediction by CHDPRS when combined with CHDPRS into CHDBioPRS. Overall, the HRs in FinnGen are smaller than in UKB for all PRS.

When comparing the CHD risk between the top 5% of the PRS distribution and the rest, we observed larger HRs for CHDBioPRS (UKB test: 4.16 [95% CI: 3.09-5.60], FinnGen: 4.04 [95% CI: 3.76-4.35]) than for CHDPRS (UKB test: 3.56 [95% CI: 2.62-4.86], FinnGen: 3.84 [95% CI: 3.56-4.14]). HRs for other percentiles along with their AUCs are listed in [Supplemental Table 6](#) for UKB Test and [Supplemental Table 7](#) for FinnGen.

We next studied the PRS in cases with early CHD onset (<55 years of age). For UKB test data, we had 325 cases (20.6% female) and for FinnGen 5,965 cases (21.4% female). For both cohorts, we observed higher HRs for early onset cases than for all cases and again we saw HRs growing when using CHDBioPRS (UKB test: 2.07 [95% CI: 1.85-2.32], FinnGen: 2.10 [95% CI: 2.04-2.16]) instead of CHDPRS (UKB test: 1.90 [95% CI: 1.69-2.13], FinnGen: 2.01 [95% CI: 1.95-2.07]) ([Figure 2](#), [Table 3](#); see also [Supplemental Table 6](#) for UKB test and [Supplemental Table 7](#) for FinnGen). Among all CHD cases, the CHDBioPRS values peaked for cases with an onset at around 40 years of age ([Supplemental Figure 3](#)).

It is known that CHD incidence and biomarker associations vary between sexes.^{3,10} Therefore, we stratified UKB data and FinnGen data by sex. In both sexes, the regularized optimal models built in UKB training data consisted of the same 8 biomarkers (HDL, LDL, TRIG, SBP, CPD, HbA_{1c}, CRP, and ApoB). Using these biomarkers, we refitted the models in each sex separately and derived corresponding scores (BioPRS and CHDBioPRS). The CHD HR had similar dynamics between CHDPRS and CHDBioPRS compared with the full cohort ([Table 3](#), [Supplemental Tables 8 to 11](#)). Among FinnGen women with onset before 60 years of age, the HRs were 1.82 (95% CI: 1.74-1.90) for CHDPRS and 1.90 (95% CI: 1.81-1.99) for CHDBioPRS ([Supplemental Table 12](#)), and for FinnGen men with onset before 50 years of age, the HRs were 2.17 (95% CI: 2.07-2.27) for CHDPRS and 2.26 (95% CI: 2.16-2.36) for CHDBioPRS ([Supplemental Table 13](#)).

We observed that CHDBioPRS provided improved prediction in terms of HR, C-index, and LRT compared with SCORE2, a prediction algorithm based only on measured risk factors, but that a combination of SCORE2 and CHDBioPRS gave the best performance ([Figure 2](#), [Supplemental Tables 18 and 19](#)). In



evaluating NRI, we found that CHDBioPRS improved NRI over CHDPRS ([Supplemental Table 16](#)) by 0.239 for UKB test, 0.301 for UKB test early onset, 0.101 for FinnGen, and 0.139 for FinnGen early onset. Finally, we also found that CHDBioPRS improved LRT compared with CHDPRS in both UKB test and FinnGen cohorts ([Supplemental Figure 4](#), [Supplemental Table 18](#)).

DISCUSSION

In an analysis of more than 600,000 participants involving 2 nationwide study cohorts, UKB and FinnGen, we showed that combining a biomarker PRS derived from 10 CHD-associated biomarkers with a standard PRS for CHD improved the prediction of CHD. This improvement in prediction was largest for early onset CHD for both men and women.

Our BioPRS compared well with the recently published multi-PRS score by Sinnott-Armstrong et al,²

which was derived from 35 UKB biomarkers. Their multi-PRS score was reported to have an HR of 1.19 (95% CI: 1.17-1.22) in FinnGen (data freeze 3) for MI, while our BioPRS achieved an HR of 1.29 (95% CI: 1.27-1.31) using the same endpoint and same covariates in the more recent and larger data freeze 7 of FinnGen. In addition, when Sinnott-Armstrong *et al* combined their multi-PRS with a standard PRS for CHD in FinnGen, they reported an HR of 1.50 (95% CI: 1.46-1.53), while our CHDBioPRS achieved an HR of 1.60 (95% CI: 1.58-1.62) in FinnGen (data freeze 7). Similarly, a MetaGRS score from Inouye *et al*,²⁶ which combined 3 genetic risk scores (GRS46K,²² another score based on 202 significant genetic variants from CARDIOGRAMplusC4D,¹⁷ and a genome-wide score based on the same GWAS¹³), reported an HR of 1.71 (95% CI: 1.68-1.73) when tested on UKB, while our CHDBioPRS yielded an HR of 1.88 (95% CI: 1.75-2.01) in our UKB test data.

There were clear differences in effect sizes between UKB and FinnGen. These differences could in part relate to differences in sample ascertainment procedures and genetic background. The UKB participants are known to be healthier than the general population,⁴⁸ and therefore the relative contribution of genetics to their disease risk may be larger, whereas the FinnGen participants are recruited through their contacts with the Finnish health care system. In addition, the FinnGen participants are on average 5 years older than the UKB participants, and the CHD case rate in FinnGen is nearly double that of UKB (10.2% vs 5.6%). Differences in performance of PRS are known to exist even between populations of European ancestry.⁴⁹ In our study, the biomarker GWAS effect sizes and LD information used in creating PRS were derived in UKB or from other non-Finnish European populations, which could lead to better predictive power of PRS in UKB compared with Finnish data.⁵⁰ For our CHDPRS, the effect sizes were taken from a large GWAS meta-analysis¹³ that may have included some of our FinnGen test samples. However, because our CHDPRS performed better in UKB than in FinnGen, we do not expect this potential overlap to have caused serious overfitting in our FinnGen test data.

Both sex-specific CHDBioPRS were constructed from the same set of 8 biomarkers (HDL, LDL, TRIG, SBP, CPD, HbA_{1c}, CRP, and ApoB) and achieved prediction improvements compared with the standard CHD PRS. Importantly, we observed fairly similar HR estimates for men and women, which is in contrast with existing CHD clinical scores, such as QRISK/QRISK2 which is known to underestimate the CHD risks in women.⁵¹⁻⁵⁴ We also observed larger HRs for

early onset cases than all cases, indicating that our PRS are also informative about age at onset.

Throughout the analyses, we saw that prediction using CHDBioPRS was statistically strongly favored over that from CHDPRS when measured by likelihood ratio. Similarly, we saw that CHDBioPRS led to higher performance compared with SCORE2,⁴² which is a predictor based on traditional risk factor measurements, and the combination of CHDBioPRS and SCORE2 performed best. Thus, when both genetic data and risk factor measurements are available, the combination of the two may be beneficial. An important caveat here is that because we were able to compute SCORE2 only on UKB, and UKB data have been used in derivation of the SCORE2 algorithm, our SCORE2 predictions in UKB data may be too optimistic.

STUDY LIMITATIONS. Disease risk prediction using multiple biomarkers and genome-wide set of genetic variants is a very high-dimensional problem and therefore adding more sparsity to the model building could improve the risk prediction. For example, one could attempt to use, for each genomic region separately, only a relevant subset of biomarkers.⁵⁵ Furthermore, genomic regions could be prioritized, for example, by curated CHD molecular pathways⁵⁶ including known lipid-associated genomic regions.⁵⁷

The present study is limited to individuals with European ancestry. Given recent discoveries about imperfect transferability of PRS between populations,^{58,59} training of an optimal CHDBioPRS for non-European ancestries would require appropriate training data from those other ancestries. This is also important because the potential to use genetic scores to identify high-risk individuals from birth could exacerbate the health differences between individuals with European ancestry and others until there is a broader inclusion of underserved ethnicities in research, particularly in multiethnic countries such as the UK and the U.S.⁶⁰

An approach similar to our BioPRS could also improve prediction of other complex diseases, such as type 2 diabetes and breast cancer, with established PRS and known heritable risk factors.⁵⁹

CONCLUSIONS

The integration of biomarker PRS improves on the standard PRS for prediction of CHD, where the gain was largest among early onset CHD cases. This study strengthens the evidence for genome-based CHD prediction and quantifies the interplay between standard CHD PRS and PRS of biomarkers associated with CHD.

ETHICS STATEMENT AND METHODS. Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected before the Finnish Biobank Act came into effect in September 2013 and the start of FinnGen in August 2017, were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by the Finnish Medicines Agency (Fimea), the national supervisory authority for welfare and health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa statement no. for the FinnGen study is HUS/990/2017.

The FinnGen study was approved by the Finnish Institute for Health and Welfare (permit nos. THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019, THL/1524/5.05.00/2020, and THL/2364/14.02/2020), the Digital and Population Data Service Agency (permit nos. VRK43431/2017-3, VRK/6909/2018-3, and VRK/4415/2019-3), the Social Insurance Institution (permit nos. KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 138/522/2019, KELA 2/522/2020, and KELA 16/522/2020), Findata (THL/2364/14.02/2020) and Statistics Finland (permit nos. TK-53-1041-17 and TK/143/07.03.00/2020 [earlier TK-53-90-20]).

The Biobank Access Decisions for FinnGen samples and data used in FinnGen data freeze 7 include THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, and BB2020_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, Auria Biobank AB17-5154 and amendment no. 1 (August 17, 2020), Biobank Borealis of Northern Finland 2017_1013, Biobank of Eastern Finland 1186/2018 and amendment 22 § /2020, Finnish Clinical Biobank Tampere MH0004 and amendments 21.02.2020 and 06.10.2020, Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001.

ACKNOWLEDGMENTS The authors acknowledge the participants and investigators of the FinnGen study. The individual contributors to FinnGen are listed in the Supplemental Appendix. The FinnGen project is funded by 2 grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie, AstraZeneca UK, Biogen MA, Bristol Myers Squibb (and Celgene Corp and Celgene International II), Genentech, Merck Sharp & Dohme,

Pfizer, GlaxoSmithKline Intellectual Property Development, Sanofi US Services, Maze Therapeutics, Janssen Biotech, Novartis, and Boehringer Ingelheim International. We acknowledge the following biobanks for delivering biobank samples to FinnGen: Auria Biobank, THL Biobank, Helsinki Biobank, Biobank Borealis of Northern Finland, Finnish Clinical Biobank Tampere, Biobank of Eastern Finland, Central Finland Biobank, Finnish Red Cross Blood Service Biobank, and Terveystalo Biobank. All Finnish biobanks are members of the Finnish Biobank Cooperative (<https://finbb.fi/>), which is the coordinator of BBMRI-ERIC (Biobanking and BioMolecular Resources Research Infrastructure-European Research Infrastructure Consortium Infrastructure) operations in Finland. The Finnish biobank data can be accessed through the Fingeniousservices (<https://site.fingenious.fi/en/>) managed by FINBB. This research was conducted using the UK Biobank Resource under application no. 22627.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

This work was supported by the Academy of Finland (grant nos. 325999 to Dr Lin, 331671 to Dr Mars, 312076 and 336825 to Dr Pirinen, and 285380 and 312062 to Dr Ripatti) and the Sigrid Juselius Foundation (to Drs Pirinen and Ripatti). Dr Lin was also supported by GEMMA (H2020-SC1-BHC-03-2018, project ID 825033). All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Jake Lin, Health Sciences, Faculty of Social Sciences, Tampere University, Arvo Ylpön katu 34, 733520 Tampere, Finland. E-mail: jake.lin@helsinki.fi.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: CHD is a leading cause of death, and there is an increased need for better genetic prediction, especially for high-risk individuals. A standard PRS for CHD has already been shown to improve prediction. A new PRS, constructed by combining the CHD PRS with PRS of statistically relevant CHD risk factors, can further improve genetic prediction, particularly for high-risk individuals. The identification of individuals with the highest risk of CHD can assist clinicians in their decision making.

TRANSLATIONAL OUTLOOK: Additional research is needed to establish the clinical benefit of the novel PRS, which includes information from multiple CHD-associated biomarkers, providing a more individualized prediction for CHD.

REFERENCES

1. Visseren FLJ, François Mach, Smulders YM, et al. 2021 ESC guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J*. 2021;42(34):3227-3337.
2. Sinnott-Armstrong Nasa, Tanigawa Y, et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat Genet*. 2021;53(2):185-194.
3. Johnston N, Jernberg T, Lagerqvist B, Siegbahn A, Wallentin L. Improved identification of patients with coronary artery disease by the use of new lipid and lipoprotein biomarkers. *Am J Cardiol*. 2006;97(5):640-645.
4. Zakynthinos E, Pappa N. Inflammatory biomarkers in coronary artery disease. *J Cardiol*. 2009;53(3):317-333.
5. Wilson P, Castelli WP, Kannel WB. Coronary risk prediction in adults (the Framingham Heart Study). *Am J Cardiol*. 1987;59(14):91G-94G.
6. d'Agostino RB Sr, Vasani RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*. 2008;117(6):743-753.
7. Mahmood S, Levy D, Vasani RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet*. 2014;383(9921):999-1008.
8. Wachira J, Stys TP. Cardiovascular disease and bridging the diagnostic gap. *S D Med*. 2013;66(9):366-369.
9. McMahan C Alex, Gidding SS, et al. Risk scores predict atherosclerotic lesions in young people. *Arch Intern Med*. 2005;165(8):883-890.
10. Aggarwal A, Srivastava S, Velmurugan M. Newer perspectives of coronary artery disease in young. *World J Cardiol*. 2016;8(12):728-734.
11. Aggarwal A, van Eyk J, Wei J, Bairey Merz CN. Sex, myocardial infarction, and the failure of risk scores in women. *J Womens Health*. 2015;24(11):859-861.
12. Garcia M, Mulvagh S, Bairey Merz CN, Buring J, Manson JoAnn. Cardiovascular disease in women: clinical perspectives. *Circ Res*. 2016;118(8):1273-1293.
13. Nikpay M, Goel A, Won Hong-Hee, et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47(10):1121-1130.
14. Van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res*. 2018;122(3):433-443.
15. Matsunaga H, Ito K, Akiyama M, et al. Trans-ethnic meta-analysis of genome-wide association studies identifies three new loci and characterizes population-specific differences for coronary artery disease. *Circ Genom Precis Med*. 2020;13(3):e002670.
16. Schunkert H, König IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43(4):333-338.
17. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45(1):25-33.
18. Nelson CP, Goel A, Butterworth AS, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet*. 2017;49(9):1385-1391.
19. Vilhjálmsson BJ, Yang J, Finucane HK, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet*. 2015;97(4):576-592.
20. Privé F, Arbel J, Vilhjálmsson BJ. Ldpred2: better, faster, stronger. *Bioinformatics*. 2020;35(22-23):5424-5431.
21. Ge T, Che C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10(1):1776.
22. Abraham G, Havulinna AS, Bhalala OG, et al. Genomic prediction of coronary heart disease. *Eur Heart J*. 2016;37(43):3267-3278.
23. Khera A, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219-1224.
24. Ripatti S, Tikkanen E, Orho-Melander M, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet*. 2010;376(9750):1393-1400.
25. Wünnemann Florian, Lo KS, et al. Validation of genome-wide polygenic risk scores for coronary artery disease in French Canadians. *Circ Genom Precis Med*. 2019;12(6):e002481.
26. Inouye M, Abraham G, Nelson CP, et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J Am Coll Cardiol*. 2018;72(16):1883-1893.
27. Munz M, Richter GM, Loos BG, et al. Genome-wide association meta-analysis of coronary artery disease and periodontitis reveals a novel shared risk locus. *Sci Rep*. 2018;8(1):13678.
28. Brewer LC, Svatikova A, Mulvagh SL. The challenges of prevention, diagnosis and treatment of ischemic heart disease in women. *Cardiovasc Drugs Ther*. 2015;29(4):355-368.
29. Mars N, Kerminen S, Yen-Chen A Feng, et al. Genome-wide risk prediction of common diseases across ancestries in one million people. *Cell Genom*. 2022;2(4):100118.
30. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: a free access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779.
31. Tamlander M, Mars N, Pirinen M, FinnGen, Widén E, Ripatti S. Integration of questionnaire-based risk factors improves polygenic risk scores for human coronary heart disease and type 2 diabetes. *Commun Biol*. 2022;5(1):158.
32. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*. 1972;18(6):499-502.
33. Loh P-R, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47(3):284-290.
34. FinnGen. FinnGen documentation of R5 release. May 11, 2021. <https://finngen.gitbook.io/documentation/v/r5/>
35. Friedman J, Hastie T, Tibshirani R, et al. glmnet: lasso and elastic-net regularized generalized linear models. <https://cran.r-project.org/web/packages/glmnet/index.html>
36. Hastie T, Qian J. Glnet vignette. September 13, 2016. https://hastie.su.domains/Papers/Glmnet_Vignette.pdf
37. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
38. Chang C. PLINK2. <https://www.cog-genomics.org/plink/2.0/>
39. Therneau TM. survival: survival analysis. <https://cran.r-project.org/web/packages/survival/index.html>
40. Fieuws S, Verbeke G. Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Stat Med*. 2004;23(20):3093-3104.
41. Verbeke G. Linear mixed models for longitudinal data. In: Verbeke G, Molenberghs G, eds. *Linear Mixed Models in Practice: A SAS-Oriented Approach*. New York: Springer; 1997:63-153.
42. SCORE2 Working Group and ESC Cardiovascular Risk Collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J*. 2021;42(25):2439-2454.
43. Maas AHM, Appelman YEA. Gender differences in coronary heart disease. *Neth Heart J*. 2010;18(12):598-602.
44. Inoue E. nricens: NRI for risk prediction models with time to event and binary response data. <https://cran.r-project.org/web/packages/nricens/index.html>
45. Studio R. R markdown. <https://rmarkdown.rstudio.com/>
46. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
47. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247(18):2543-2546.
48. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol*. 2017;186(9):1026-1034.
49. Damian Gola, Erdmann J, Läll K, Mägi Reedik, et al. Population bias in polygenic risk prediction

models for coronary artery disease. *Circ Genom Precis Med.* 2020;13(6):e002932.

50. Kerminen S, Martin AR, Koskela J, et al. Geographic variation and bias in the polygenic scores of complex diseases and traits in Finland. *Am J Hum Genet.* 2019;104(6):1169-1181.

51. Saar A, Lall K, Alver M, et al. Estimating the performance of three cardiovascular disease risk scores: the Estonian Biobank cohort study. *J Epidemiol Community Health.* 2019;73:272-277.

52. Sedlak T, Herscovici R, Cook-Wiens G, et al. Predicted versus observed major adverse cardiac event risk in women with evidence of ischemia and no obstructive coronary artery disease: a report from WISE (Women's Ischemia Syndrome Evaluation). *J Am Heart Assoc.* 2020;9(7):e013234.

53. Woodward M. Cardiovascular disease and the female disadvantage. *Int J Environ Res Public Health.* 2019;16(7):1165.

54. Bots SH, Peters SAE, Woodward M. Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of ageing between 1980 and 2010. *BMJ Global Health.* 2017;2:e000298.

55. Lin J, Tabassum R, Ripatti S, Pirinen M. MetaPhat: detecting and decomposing multivariate associations from univariate genome-wide association statistics. *Front Genet.* 2020;11:431.

56. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MsigDB) hallmark gene set collection. *Cell Syst.* 2015;1(6):417-425.

57. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet.* 2013;45(11):1274-1283.

58. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current poly-

genic risk scores may exacerbate health disparities. *Nat Genet.* 2019;51(4):584-591.

59. Mars N, Koskela JT, Ripatti P, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med.* 2020;26(4):549-557.

60. Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. *Cell.* 2018;177(1):26-31.

KEY WORDS biomarkers, coronary heart disease, genomics, GWAS, polygenic risk scores

APPENDIX For supplemental figures and tables and a list of the contributors of FinnGen, please see the online version of this paper.