Integration of biomarker polygenic risk score improves prediction of coronary heart disease in UK Biobank and FinnGen

Jake Lin1*, Nina Mars1, Yu Fu1, Pietari Ripatti1, Tuomo Kiiskinen1, FinnGen, Taru Tukiainen1, Samuli Ripatti1,2,3, Matti Pirinen1,2,4*

1. Institute for Molecular Medicine Finland FIMM, Helsinki Institute of Life Science HiLIFE, University of Helsinki, Helsinki, Finland.
2. Department of Public Health, University of Helsinki, Helsinki, Finland.
3. Massachusetts General Hospital & Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA.
4. Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland.

Abstract

Background

In addition to age and sex, also smoking history and levels of blood pressure, cholesterol, lipoproteins and inflammation are established biomarkers for coronary heart disease (CHD). As standard polygenic risk scores (PRS) have recently proven successful for CHD prediction, it remains of high interest to determine how a combined PRS of biomarkers (BioPRS) constructed from statistically relevant biomarkers can further improve genetic prediction of CHD.

Methods

We developed CHDBioPRS, which combines BioPRS with PRS of CHD, via regularized regression in UK Biobank (UKB) training data (n = 208,010). The resulting CHDBioPRS was tested on an independent UK Biobank subset (n = 25,765) and on the FinnGen study (n = 306,287).

Results

We observed a consistent pattern across all data sets where BioPRS was clearly predictive of CHD and improved standard PRS for CHD when the two were combined. In UKB test data, CHDPRS had a hazard ratio (HR) of 1.78 (95% confidence interval 1.67-1.91, area under the curve (AUC) 0.808) and CHDBioPRS had a HR of 1.88 (1.75-2.01, AUC 0.811) per one standard deviation of PRS. In FinnGen data, HR of CHDPRS was 1.57 (1.55-1.60, AUC 0.752) and HR of CHDBioPRS was 1.60 (1.58-1.62, AUC 0.755). We observed larger effects of CHDBioPRS in subsets of early onset cases with HR of 2.07 (1.85-2.32, AUC 0.790) in UKB test data and of 2.10 (2.04-2.16, AUC 0.791) in FinnGen. Results were similar when stratified by sex.

Conclusions

Integration of biomarker based BioPRS improved on the standard PRS for CHD and the gain was largest with early onset CHD cases. These findings highlight the benefit of enriching polygenic risk prediction of CHD with the genetics of associated biomarkers.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
BACKGROUND

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 (Visseren et al. 2021; Sinnott-Armstrong et al. 2021; Johnston et al. 2006; Zakynthinos and Pappa 2009). These risk factors are also used in clinical risk calculators to evaluate preventive therapies and strategies. While clinical risk scores enable identification of some individuals at high risk (Wilson, Castelli, and Kannel 1987; D’Agostino et al. 2008; Mahmood et al. 2014), a large proportion of CHD cases are not detected by these scores and the utility of clinical scores is limited for young adults (Wachira and Stys 2013; Johnston et al. 2006; McMahan et al. 2005; Aggarwal, Srivastav; Berry et al. 2007) and for women (Aggarwal et al. 2015; Garcia et al. 2016).

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 (Nikpay et al. 2015; van der Harst and Verweij 2018; Matsunaga et al. 2020; Schunkert et al. 2011; CARDIoGRAMplusC4D Consortium et al. 2013; Nelson et al. 2017). These genetic discoveries together with the introduction of sophisticated statistical tools that incorporate linkage disequilibrium information, have greatly advanced risk prediction (Vilhjálmsson et al. 2015; Privé, Arbel, and Vilhjálmsson 2020; Ge et al. 2019). Particularly, several studies have shown that inclusion of polygenic risk scores (PRSes) improve CHD prediction and identification of high-risk groups (Abraham et al. 2016; Nikpay et al. 2015; Khera et al. 2018; Ripatti et al. 2010). Because PRSes 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 PRSes for common diseases conducted by Khera et al. (Khera et al. 2018) showed that a sizable portion of the population carry a polygenic CHD risk equivalent to known monogenic mutations conferring several-fold increased risk. This PRS, comprising more than 6 million SNPs, was generated using LDpred (Vilhjálmsson et al. 2015) and has proven effective in
validation sets across multiple populations while also performing favorably compared to other polygenic risk scores (Wünnemann et al. 2019) composed of smaller numbers of variants.

As PRSes have proven successful for CHD prediction, it remains of high interest to systematically determine how a combined biomarker score (BioPRS) constructed from biomarkers associated with CHD can improve upon the established CHDPRS. Recently, multi-PRS models, using 35 PRSes from blood and urine biomarkers, have been shown to improve genetic risk prediction of common disease such as type 2 diabetes and gout (Sinnott-Armstrong et al. 2021). While other existing studies have focused on combining several GWAS of CHD (Inouye et al. 2018; Munz et al. 2018; Nikpay et al. 2015), our focus is on the combination of effects of known CHD associated biomarkers into a single PRS and its integration with CHDPRS. Furthermore, as 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 (Brewer, Svatikova, and Mulvagh 2015). Another important goal is to predict a subgroup of CHD cases with an early onset of the disease (Mars et al. 2020).

**METHODS**

The workflow of the study is in Figure 1.

**UKB data:** The design of UKB and the background of its participants have been reported previously (Sudlow et al. 2015; Abraham et al. 2016). 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 individuals from different ancestry backgrounds in our GWAS. We made use of the following four 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 below) 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 consistent with previous studies (Khera et al. 2018; Tamlander et al. 2022), CHD cases were defined as having a heart attack diagnosed by a doctor (field 6150) or self-reported as a non-cancer illness (field 20002) or operations including PTCA and CABG (field 20004). In addition, coronary revascularization was assigned based on OPCS-4 (field 41272) 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 hospital episode statistics (fields 41202 (ICD-10) and 41203 (ICD-9) ) were defined with ICD-10 I21, I22, I23, I24.1 and I25.2 and ICD-9 codes of 410, 411 and 412. We further merged the outcomes from the UKB MI algorithmically-defined outcomes and dates (fields 42000 and 42001). 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 while 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 associated with CHD (Table 2 and Supplementary Table S1): HDL cholesterol (field 30760), LDL direct (30780), Triglycerides (TRIG, 30870), ApoA1 (30630), ApoB (30640), diastolic blood pressure (DBP, 4079), systolic blood pressure (SBP, 4080), glycedated hemoglobin (HbA1c, 30750), glucose (30740), C-reactive protein (CRP, 30710), creatinine (CREA, 30700), lipoprotein(a) (LPA, 30790) doctor diagnosed diabetes (2443), body mass index (BMI, 21001) and cigarettes per day (CPD, 3456). In addition, total cholesterol (TC) was calculated from the Friedewald formula (Friedewald, Levy, and Fredrickson 1972) as HDL + LDL + TG/2.2 in units of mmol/L. Following the previous biomarker study by Sinnott-Armstrong et al. (Sinnott-Armstrong et al. 2021), statin usage is identified from treatment/medication (field 20003, 13 drugs: 1141146234, atorvastatin; 1141192414, crestor 10mg tablet; 1140910632, eptastatin; 1140888594, fluvastatin; 1140864592, lesco 20 mg capsule; 1141146138, lipitor 10mg tablet; 1140861970, lipostat 10mg tablet; 114088648, pravastatin; 1141192410, rosuvastatin; 1141188146, simvador 10mg tablet; 1140861958, simvastatin; 1140881748, zocor 10mg tablet; 1141200040, zocor heart-pro 10mg tablet). Adjustment for statin usage (Sinnott-Armstrong et al. 2021) 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), LPA (1.102), GLUC (1.029), CRP (1.230), CREA (1.058) and HbA1c (1.042). We also adjusted the subjects taking blood pressure reduction drugs (fields 6153 and 6177) by adding +15 to SBP and +10 to DBP. The statin adjusted values of these biomarkers for the UKB training set are shown in Table 2.
Genome-wide association study (GWAS), using BOLT-LMM v2.3.2 (Loh et al. 2015) 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$^2$ and the top 10 principal components of genetic structure from the biomarkers and then applied rank-based inverse-normal transformation to the residuals.

**FinnGen:** The design of the Finnish FinnGen (FinnGen 2021) project and participant backgrounds are shown in Table 1. The FinnGen test cohort contained 321,302 FinnGen data freeze 7 (DF7) 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. Subsequently, we removed the 2,989 angina only cases from FinnGen, in addition to removal of 6,109 participants less than 16 years of age at enrollment.

**Biomarker model:** Using as predictors the 16 CHD associated biomarkers for the UKB training data and incident CHD as outcome (and excluding prevalent CHD cases), we applied elastic net Cox regression as implemented in glmnet (Friedman et al. 2021; Hastie and Qian 2014) with alpha = 0.50 and using 20-fold cross validation. We chose the optimal model by glmnet’s “lambda.1se” criterion and identified 10 biomarkers (Table 2). As shown in the Supplementary Figure 1, an alpha = 0.25 produced the same optimal set as 0.50 setting while alpha = 0.75 further excluded ApoA1. The identified 10-biomarker model from lambda.1se was selected for BioPRS construction.

**Biomarker PRSes:** PRS-CS (Ge et al. 2019) was run on each of the GWAS summary results of the selected 10 biomarkers. To account for linkage disequilibrium (LD), we employed the 1000 Genomes project’s (Nikpay et al. 2015; 1000 Genomes Project Consortium et al. 2015) 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 using PLINK2.0 (Purcell and Chang, 2019).

**Weights of biomarker PRSes on CHD:** The PRSes 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:

\[
(1) \quad h(t) = h_0(t) \cdot \exp \left( b_{hdp}x_{hdp} + b_{ld}x_{ld} + b_{tg}x_{tg} + b_{apo}x_{apo} + b_{apo}x_{apo} + b_{cp}x_{cp} + b_{cre}x_{cre} + b_{cpd}x_{cpd} + b_{sby}x_{sby} + b_{x}x \right)
\]
Here \( h_0(t) \) 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’ library (Therneau 2021) of R software to estimate the \( b_{\text{biomarkerPRS}} \) coefficients as previously recommended (Fieuws and Verbeke 2004; Verbeke 1997).

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

\[
(2) \text{BioPRS} = \text{standardize} \left( \sum_{i=1}^{10} b_i \cdot PRS_i \right)
\]

**CHDPRS:** We generated a PRS for CHD (named CHDPRS) by applying PRS-CS (Ge et al. 2019) to CHD GWAS reported by Nikpay et al. (2015) using European panel from the 1000 Genomes Project (1000 Genomes Project Consortium et al. 2015) for LD reference. Our CHDPRS contained 1,087,715 SNPs. Additionally, we compared our CHDPRS to “Khera PRS” that is the PRS for CHD generated using LDpred by Khera et al. (2018) based on the same GWAS (Nikpay et al. 2015) that we used to generate our CHDPRS.

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

\[
(3) h(t) = h_0(t) \cdot \exp(c_{\text{CHDPRS}}x_{\text{CHDPRS}} + c_{\text{BioPRS}}x_{\text{BioPRS}})
\]

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

\[
(4) \text{CHDBioPRS} = \text{standardize}(c_{\text{CHDPRS}}x_{\text{CHDPRS}} + c_{\text{BioPRS}}x_{\text{BioPRS}})
\]

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).

**Early onset:** We identified all individuals with early CHD onset (< 55 years) 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.
**Statistical analysis:** All scores were standardized to have a mean of 0 and a variance of 1. Hazard ratio (HR) and its 95% confidence interval were computed using ‘coxph’ from the survival package. In Cox regression, age was used as the time scale. When evaluating PRSes, 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 controls were censored at the upper limit of the early onset age.

C-Index (Harrell et al. 1982), a metric for prediction concordance, is obtained from the ‘survival’ (Therneau 2021) package in R software and Area under the curve (AUC) were used to assess model discrimination. AUC and its confidence interval were obtained from logistic regression (‘glm’ function) and ‘pROC’ package (Robin et al. 2011) in R software. We also computed how much each biomarker PRS explains of the variance of the corresponding biomarker by using the adjusted R² measure as reported by ‘lm’ function in R software, for the linear model where the biomarker PRS was the only predictor in the model. This analysis was done in the UKB test data.

**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 of 16,658 CHD cases (22.6% female). We further split this data into disjoint 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 (56.4% women; 10.4% CHD cases of whom 27.6% women) Finnish individuals. A combined PRS of CHD biomarkers (BioPRS) was derived across PRSes of 10 biomarkers (HDL, LDL, TRIG, Total cholesterol, systolic blood pressure, cigarettes per day, HbA1c, C-Reactive protein, creatinine, ApoB, ApoA1) by combining the individual biomarker PRSes generated with PRS-CS on 1,106,191 SNPs. This BioPRS was further integrated with the CHDPRS constructed from 6,630,150 variants from a GWAS (Nikpay et al. 2015) involving 184,305 participants of European ancestry to yield CHDBioPRS. We also compared our CHDPRS to the Khera PRS (Khera et al. 2018) and saw that our CHDPRS gave slightly more predictive result in UKB training data (HR of 1.64, 95%CI (1.61-1.67) compared to 1.62 (1.59-1.64) (Supplementary Table S2)). CHDBioPRS is approximately normally distributed.
and, on average, higher in CHD cases relative to the controls (mean difference 0.518 SD, with 95% CI (0.490,0.546); Supplementary Figure S2).

Table 3 shows the results from UKB data sets and FinnGen. In UKB, results from the training set are similar to results from validation and test sets, which suggests that the model is not overfitting in the training data. In all three UKB data sets, we observe the same pattern where BioPRS itself is clearly predictive of CHD (HR estimates per standard deviation vary between 1.42-1.45), CHDPRS on its own is more predictive than BioPRS (HR estimates vary between 1.62-1.78) and CHDBioPRS is the most predictive (HRs vary between 1.73-1.88). For C-Index, Z-scores and AUC metrics see Supplementary Tables S2, S3 and S4.

When the scores are applied in FinnGen (Supplementary Tables S5), we repeat the pattern 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 PRSes.

By 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 (3.09-5.60), FinnGen 4.04 (3.76-4.35)) than for CHDPRS (UKB Test 3.56 (2.62-4.86), FinnGen 3.84 (3.56-4.14), HRs for other percentiles along with their AUCs are listed in Supplementary Tables S6 for UKB Test and S7 for FinnGen).

We next studied the PRSes in cases with early CHD onset (< 55 years). For UKB test data, we had 325 cases (females 20.6%) and for FinnGen, 5965 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 (1.85-2.32), FinnGen 2.10 (2.04-2.16)) instead of CHDPRS (UKB test 1.90 (1.69-2.13, FinnGen 2.01 (1.95-2.07) Table 3; see also Supplementary Tables S6 for UKB Test and S7 for FinnGen). Among all CHD cases, the CHDBioPRS values peaked for cases with an onset at around 40 years of age (Supplementary Figure S3).

Among FinnGen women with onset below 60 years of age, the HRs were 1.82 (1.74-1.90) for CHDPRS and 1.90 (1.81-1.99) for CHDBioPRS (Supplementary Table S12) and for FinnGen males with onset below 50 years of age, the HRs were 2.17 (2.07-2.27) for CHDPRS and 2.26 (2.16-2.36) for CHDBioPRS (Supplementary Table S13).
DISCUSSION

In an analysis of more than 600,000 participants involving two nationwide study cohorts, UK biobank and FinnGen, we showed that combining a biomarker polygenic risk score derived from 10 CHD associated biomarkers with a standard PRS for CHD improved on prediction of CHD. This improvement in prediction was largest for the early-onset CHD for both men and women. Our BioPRS compared well with the recently published multi-PRS score by Sinnott-Armstrong et al. (Sinnott-Armstrong et al. 2021) that was derived from 35 UKB biomarkers. Their multi-PRS score was reported to have a HR of 1.19 (1.17-1.22) in FinnGen (data freeze 3) for myocardial infarction (Sinnott-Armstrong et al. 2021) while our BioPRS achieved HR of 1.29 (1.27-1.31) using the same endpoint and same covariates in a more recent and larger data freeze 7 of FinnGen. Additionally, when Sinnott-Armstrong et al. combined their multi-PRS with a standard PRS for CHD in FinnGen, they reported a HR of 1.50 (1.46-1.53) (Sinnott-Armstrong et al. 2021), while our CHDBioPRS achieved a HR of 1.60 (1.58-1.62) in FinnGen (data freeze 7). Similarly, a MetaGRS score from Inouye et al. (Inouye et al. 2018), that combined three genetic risk scores (GRS46K (Abraham et al. 2016), another score based on 202 significant genetic variants from CARDIOGRAMplusC4D (Nikpay et al. 2015) and a genome-wide score based on the same GWAS (Nikpay et al. 2015)), reported a HR of 1.71 (1.68-1.73) when tested on UKB while our CHDBioPRS yielded a HR of 1.88 (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 (Fry et al. 2017) and therefore the relative contribution of genetics to their disease risk may be larger, whereas the FinnGen participants have been recruited through their contacts with the Finnish healthcare system. Additionally, the FinnGen participants are on average 5 years older compared to the UK biobank participants and the CHD case rate in FinnGen is nearly double compared to UK biobank (10.2% vs. 5.6%). The biomarker GWAS effect sizes and linkage disequilibrium information used in creating PRS were derived in UKB or from other non-Finnish European populations, which could lead to better predictive power of PRSes in UKB compared to Finnish data (Kerminen et al. 2019). For our CHDPRS, the effect sizes were taken from a large GWAS meta-analysis (Nikpay et al. 2015) that may have included some of our FinnGen test samples. However, since 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 CHDBioPRSes were constructed from the same set of 8 biomarkers (HDL, LDL, triglycerides, systolic blood pressure, cigarettes per day, HbA1c, CRP, ApoB) and achieved prediction improvements compared to the standard CHD PRS. Importantly, we observed fairly similar hazard ratio estimates for men and women, which is in contrast with existing CHD clinical scores, such as QRISK/QRISK2 that is known to underestimate the CHD risks in women (Saar et al. 2019; Sedak et al. 2020; Woodward 2019). We also observed larger hazard ratios for early onset cases than all cases indicating that our PRSes are also informative about age at onset.

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 (Lin et al. 2020). Furthermore, genomic regions could be prioritized, for example, by curated CHD molecular pathways (Liberzon et al. 2015) including known lipid-associated genomic regions (Willer et al. 2013).

Our study is limited to individuals with European ancestry. Given recent discoveries about imperfect transferability of PRS between populations (Martin et al. 2019; Mars et al. 2022), training of an optimal CHDBioPRS for non-European ancestries would require appropriate training data from other ancestries. This is also important since the potential to use genetic scores to identify high risk individuals from birth could exacerbate the health differences between individuals with European ancestry and the others until a broader inclusion of underserved ethnicities in research become reality, particularly in multi-ethnic countries such as the UK and the US (Sirugo, Williams, and Tishkoff 2019).

An approach similar to our BioPRS could also improve prediction of other complex diseases, such as Type 2 diabetes and breast cancer, with established PRSes and known, heritable risk factors (Mars et al. 2020).

CONCLUSION

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

Table 1: Sample characteristics in UK Biobank (UKB) and FinnGen. Values are either mean ± standard deviation or count (percentage). For incident CHD cases, age at onset is after the blood sample was taken.

<table>
<thead>
<tr>
<th></th>
<th>UKB</th>
<th>UKB Female</th>
<th>UKB Male</th>
<th>FinnGen</th>
<th>FinnGen Female</th>
<th>FinnGen Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>297158</td>
<td>158382</td>
<td>138776</td>
<td>300496</td>
<td>169508</td>
<td>130988</td>
</tr>
<tr>
<td>Enrollment age</td>
<td>56.9 ± 7.99</td>
<td>56.7 ± 7.90</td>
<td>57.12 ± 8.09</td>
<td>53.88 ± 16.63</td>
<td>52.35 ± 16.96</td>
<td>55.86 ± 15.74</td>
</tr>
<tr>
<td>Incident CHD cases (%)</td>
<td>4866 (1.66)</td>
<td>1321 (0.95)</td>
<td>4045 (2.91)</td>
<td>13623 (4.53)</td>
<td>3652 (2.15)</td>
<td>9971 (7.61)</td>
</tr>
<tr>
<td>Age at onset for incident cases</td>
<td>65.13 ± 6.90</td>
<td>66.01 ± 6.76</td>
<td>64.81 ± 6.92</td>
<td>70.29 ± 10.80</td>
<td>72.56 ± 12.16</td>
<td>69.46 ± 10.13</td>
</tr>
<tr>
<td>Follow-up years</td>
<td>4.85 ± 2.37</td>
<td>5.13 ± 2.35</td>
<td>4.74 ± 2.37</td>
<td>10.43 ± 8.96</td>
<td>9.22 ± 8.72</td>
<td>10.89 ± 9.01</td>
</tr>
<tr>
<td>All CHD cases</td>
<td>16658 (5.60)</td>
<td>3762 (2.38)</td>
<td>12896 (9.30)</td>
<td>30729 (10.22)</td>
<td>8486 (5.00)</td>
<td>22243 (16.98)</td>
</tr>
<tr>
<td>Age at onset</td>
<td>60.36 ± 10.19</td>
<td>61.92 ± 10.76</td>
<td>59.90 ± 9.98</td>
<td>65.37 ± 11.82</td>
<td>67.97 ± 12.50</td>
<td>64.37 ± 11.39</td>
</tr>
</tbody>
</table>

Table 2: UKB Biomarkers mean levels after adjustment for statins. Statin adjustment is done either by division (/) or addition (+) by the value given in the last column for statin users (16.2%).

<table>
<thead>
<tr>
<th>Biomarker (units)</th>
<th>Controls (females)</th>
<th>Incident CHD cases (females)</th>
<th>Statin adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ApoA1* (g/L)</td>
<td>1.52 (1.61)</td>
<td>1.42 (1.55)</td>
<td>/ 1.065364</td>
</tr>
<tr>
<td>ApoB* (g/L)</td>
<td>1.08 (1.07)</td>
<td>1.17 (1.17)</td>
<td>/ 0.721928</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>27.31 (26.99)</td>
<td>28.40 (28.05)</td>
<td></td>
</tr>
<tr>
<td>Cigarettes per day* (integer)</td>
<td>1.05 (0.92)</td>
<td>2.28 (2.52)</td>
<td></td>
</tr>
<tr>
<td>Creatinine* (umol/L)</td>
<td>71.60 (63.98)</td>
<td>75.98 (63.65)</td>
<td>/ 1.0580718</td>
</tr>
<tr>
<td>C-reactive protein* (mg/L)</td>
<td>2.48 (2.61)</td>
<td>3.12 (3.82)</td>
<td>/ 1.2300281</td>
</tr>
<tr>
<td>Diastolic BP* (mmHg)</td>
<td>84.12 (82.24)</td>
<td>88.15 (86.07)</td>
<td>+ 10 (BP meds)</td>
</tr>
<tr>
<td>Glucose (mmol/L)</td>
<td>5.07 (5.04)</td>
<td>5.29 (5.24)</td>
<td>/ 1.028824</td>
</tr>
<tr>
<td>HbA1c* (mmol/mol)</td>
<td>35.54 (35.41)</td>
<td>38.0 (37.53)</td>
<td>/ 1.0418022</td>
</tr>
<tr>
<td>HDL* (mmol/L)</td>
<td>1.46 (1.60)</td>
<td>1.29 (1.48)</td>
<td>/ 1.053</td>
</tr>
<tr>
<td>LipoA (mmol/L)</td>
<td>34.25 (35.17)</td>
<td>38.59 (37.81)</td>
<td>/ 1.101954</td>
</tr>
<tr>
<td>LDL* (mmol/L)</td>
<td>3.78 (3.80)</td>
<td>4.07 (4.17)</td>
<td>/ 0.684</td>
</tr>
<tr>
<td>Systolic BP* (mmHg)</td>
<td>140.8 (137.9)</td>
<td>150.7 (149.3)</td>
<td>+ 15 (BP meds)</td>
</tr>
<tr>
<td>Total Chol (mmol/L)</td>
<td>6.06 (6.16)</td>
<td>6.37 (6.62)</td>
<td></td>
</tr>
<tr>
<td>Triglycerides* (mmol/L)</td>
<td>1.78 (1.58)</td>
<td>2.21 (1.96)</td>
<td>/ 0.874</td>
</tr>
<tr>
<td>T2D (boolean)</td>
<td>4.1% (3.1%)</td>
<td>9.2% (7.3%)</td>
<td></td>
</tr>
</tbody>
</table>
*selected in optimal regularized model

Table 3: Hazard ratios (with 95% confidence intervals) from Cox regression model of three different polygenic risk scores (PRS) on incident CHD. BioPRS is combination of PRSes of 10 CHD related biomarkers, CHDPRS is a standard PRS for CHD, CHDBioPRS combines BioPRS and CHDPRS. Early onset was defined as CHD below 55 years of age. C-index, AUC and P-values are in Supplementary Tables S2-S7.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>BioPRS</th>
<th>CHDPRS</th>
<th>CHDBioPRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKB Training</td>
<td>1.45 (1.42-1.48)</td>
<td>1.64 (1.61-1.67)</td>
<td>1.76 (1.72-1.79)</td>
</tr>
<tr>
<td>UKB Validation</td>
<td>1.42 (1.38-1.46)</td>
<td>1.62 (1.57-1.67)</td>
<td>1.73 (1.68-1.78)</td>
</tr>
<tr>
<td>UKB Test</td>
<td>1.45 (1.36-1.55)</td>
<td>1.78 (1.67-1.91)</td>
<td>1.88 (1.76-2.01)</td>
</tr>
<tr>
<td>UKB Test females</td>
<td>1.53 (1.34-1.75)</td>
<td>1.72 (1.50-1.98)</td>
<td>1.86 (1.62-2.14)</td>
</tr>
<tr>
<td>UKB Test males</td>
<td>1.42 (1.32-1.53)</td>
<td>1.72 (1.60-1.85)</td>
<td>1.84 (1.71-1.98)</td>
</tr>
<tr>
<td>UKB Test early onset</td>
<td>1.60 (1.43-1.78)</td>
<td>1.90 (1.69-2.13)</td>
<td>2.07 (1.85-2.32)</td>
</tr>
<tr>
<td>FinnGen</td>
<td>1.27 (1.26-1.29)</td>
<td>1.57 (1.55-1.60)</td>
<td>1.60 (1.58-1.62)</td>
</tr>
<tr>
<td>FinnGen females</td>
<td>1.27 (1.24-1.30)</td>
<td>1.53 (1.50-1.57)</td>
<td>1.56 (1.53-1.60)</td>
</tr>
<tr>
<td>FinnGen males</td>
<td>1.27 (1.25-1.29)</td>
<td>1.58 (1.55-1.60)</td>
<td>1.61 (1.59-1.63)</td>
</tr>
<tr>
<td>FinnGen early onset</td>
<td>1.51 (1.47-1.55)</td>
<td>2.01 (1.95-2.07)</td>
<td>2.10 (2.04-2.16)</td>
</tr>
</tbody>
</table>

Supplementary tables:
https://docs.google.com/spreadsheets/d/1AR8Io_lrzcCJoXz_ETZuSLB5waXD6iM8/edit#gid=2034380250
Ethics statement and materials & methods:

Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea (Finnish Medicines Agency), 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 (HUS) statement number for the FinnGen study is Nr HUS/990/2017.


Acknowledgements:

This work was supported by the Academy of Finland (grant nos. 325999 for JL, 331671 for NM, 312076 and 336825 for MP, 285380 and 312062 for SR), by Sigrid Juselius Foundation (MP and SR).


This research has been conducted using the UK Biobank Resource under Application Number 22627.

**Contributors of FinnGen**

**Steering Committee**

Aarno Palotie  Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland  
Mark Daly  Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

**Pharmaceutical companies**

Bridget Riley-Gills  Abbvie, Chicago, IL, United States  
Howard Jacob  Abbvie, Chicago, IL, United States  
Dirk Paul  Astra Zeneca, Cambridge, United Kingdom  
Heiko Runz  Biogen, Cambridge, MA, United States  
Sally John  Biogen, Cambridge, MA, United States  
George Okafo  Boehringer Ingelheim, Ingelheim am Rhein, Germany  
Nathan Lawless  Boehringer Ingelheim, Ingelheim am Rhein, Germany  
Robert Plenge  Celgene, Summit, NJ, United States/Bristol Myers Squibb, New York, NY, United States  
Joseph Maranville  Celgene, Summit, NJ, United States/Bristol Myers Squibb, New York, NY, United States  
Mark McCarthy  Genentech, San Francisco, CA, United States  
Julie Hunkapiller  Genentech, San Francisco, CA, United States  
Meg Ehm  GlaxoSmithKline, Brentford, United Kingdom  
Kirsi Auro  GlaxoSmithKline, Brentford, United Kingdom  
Simonne Longerich  Merck, Kenilworth, NJ, United States  
Caroline Fox  Merck, Kenilworth, NJ, United States  
Anders Mälarstig  Pfizer, New York, NY, United States  
Katherine Klinger  Sanofi, Paris, France  
Deepak Raipal  Sanofi, Paris, France  
Eric Green  Maze Therapeutics, San Francisco, CA, United States
Robert Graham                    Maze Therapeutics, San Francisco, CA, United States
Robert Yang                     Janssen Biotech, Beerse, Belgium
Chris O’Donnell                 Novartis, Basel, Switzerland

University of Helsinki & Biobanks

Tomi Mäkelä                      HiLIFE, University of Helsinki, Finland, Finland
Jaakko Kaprio                   Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland, Finland
Petri Virolainen                Auria Biobank / University of Turku / Hospital District of Southwest Finland, Turku, Finland
Antti Hakanen                   Auria Biobank / University of Turku / Hospital District of Southwest Finland, Turku, Finland
Terhi Kilpi                     THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Markus Perola                   THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Jukka Partanen                   Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland
Anne Pitkäranta                  Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki
Juhani Junnila                   Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Raisa Serpi                     Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Tarja Laitinen                   Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District, Tampere, Finland
Veli-Matti Kosma                Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland
Jari Laukkanen                  Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland
Marco Hautalahti               FINBB – Finnish biobank cooperative

Other Experts/ Non-Voting Members

Outi Tuovila                   Business Finland, Helsinki, Finland
Scientific Committee

Pharmaceutical companies

Jeffrey Waring  
Abbvie, Chicago, IL, United States

Bridget Riley-Gillis  
Abbvie, Chicago, IL, United States

Ioanna Tachmazidou  
AstraZeneca, Cambridge, United Kingdom

Chia-Yen Chen  
Biogen, Cambridge, MA, United States

Heiko Runz  
Biogen, Cambridge, MA, United States

Zhihao Ding  
Boehringer Ingelheim, Ingelheim am Rhein, Germany

Marc Jung  
Boehringer Ingelheim, Ingelheim am Rhein, Germany

Shameek Biswas  
Celgene, Summit, NJ, United States/Bristol Myers Squibb, New York, NY, United States

Sarah Pendergrass  
Genentech, San Francisco, CA, United States

Julie Hunkapiller  
Genentech, San Francisco, CA, United States

Meg Ehm  
GlaxoSmithKline, Brentford, United Kingdom

David Pulford  
GlaxoSmithKline, Brentford, United Kingdom

Neha Raghavan  
Merck, Kenilworth, NJ, United States

Adriana Huertas-Vazquez  
Merck, Kenilworth, NJ, United States

Jae-Hoon Sul  
Merck, Kenilworth, NJ, United States

Anders Mälarstig  
Pfizer, New York, NY, United States

Xinli Hu  
Pfizer, New York, NY, United States

Katherine Klinger  
Sanofi, Paris, France

Matthias Gossel  
Sanofi, Paris, France

Robert Graham  
Maze Therapeutics, San Francisco, CA, United States

Eric Green  
Maze Therapeutics, San Francisco, CA, United States

Sahar Mozaffari  
Maze Therapeutics, San Francisco, CA, United States

Dawn Waterworth  
Janssen Research & Development, LLC, Spring House, PA, United States
Nicole Renaud  
Novartis, Basel, Switzerland

Ma´en Obeidat  
Novartis, Basel, Switzerland

**University of Helsinki & Biobanks**

Samuli Ripatti  
Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland

Johanna Schleutker  
Auria Biobank / Univ. of Turku / Hospital District of Southwest Finland, Turku, Finland

Markus Perola  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Mikko Arvas  
Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland

Olli Carpén  
Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki

Reetta Hinttala  
Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland

Johannes Kettunen  
Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland

Arto Mannermaa  
Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland

Katriina Aalto-Setälä  
Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District, Tampere, Finland

Mika Kähönen  
Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District, Tampere, Finland

Jari Laukkananen  
Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland

Johanna Mäkelä  
FINBB – Finnish biobank cooperative

**Clinical Groups**

**Neurology Group**

Reetta Kälviäinen  
Northern Savo Hospital District, Kuopio, Finland

Valterri Julkunen  
Northern Savo Hospital District, Kuopio, Finland

Hilkka Soininen  
Northern Savo Hospital District, Kuopio, Finland

Anne Remes  
Northern Ostrobothnia Hospital District, Oulu, Finland
Mikko Hiltunen  Northern Savo Hospital District, Kuopio, Finland
Jukka Peltola  Pirkanmaa Hospital District, Tampere, Finland
Minna Raivio  Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Pentti Tienari  Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Juha Rinne  Hospital District of Southwest Finland, Turku, Finland
Roosa Kallionpää  Hospital District of Southwest Finland, Turku, Finland
Juulia Partanen  Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Ali Abbasi  Abbvie, Chicago, IL, United States
Adam Ziemann  Abbvie, Chicago, IL, United States
Jeffrey Waring  Abbvie, Chicago, IL, United States
Sahar Esmaeeli  Abbvie, Chicago, IL, United States
Nizar Smaoui  Abbvie, Chicago, IL, United States
Anne Lehtonen  Abbvie, Chicago, IL, United States
Susan Eaton  Biogen, Cambridge, MA, United States
Heiko Runz  Biogen, Cambridge, MA, United States
Sanni Lahdenperä  Biogen, Cambridge, MA, United States
Janet van Adelsberg  Celgene, Summit, NJ, United States/ Bristol Myers Squibb, New York, NY, United States
Shameek Biswas  Celgene, Summit, NJ, United States/ Bristol Myers Squibb, New York, NY, United States
Julie Hunkapiller  Genentech, San Francisco, CA, United States
Natalie Bowers  Genentech, San Francisco, CA, United States
Edmond Teng  Genentech, San Francisco, CA, United States
Sarah Pendergrass  Genentech, San Francisco, CA, United States
Fanli Xu  GlaxoSmithKline, Brentford, United Kingdom
David Pulford  GlaxoSmithKline, Brentford, United Kingdom
Kirsi Auro  GlaxoSmithKline, Brentford, United Kingdom
Laura Addis  GlaxoSmithKline, Brentford, United Kingdom
John Eicher  GlaxoSmithKline, Brentford, United Kingdom
Qingqin S Li  Janssen Research & Development, LLC, Titusville, NJ 08560, United States
Karen He  Janssen Research & Development, LLC, Spring House, PA, United States
Ekaterina Khramtsova  Janssen Research & Development, LLC, Spring House, PA, United States
Gastroenterology Group

Martti Färkkilä Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Jukka Koskela Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Sampsa Pikkarainen Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Airi Jussila Pirkanmaa Hospital District, Tampere, Finland
Katri Kaukinen Pirkanmaa Hospital District, Tampere, Finland
Timo Blomster Northern Ostrobothnia Hospital District, Oulu, Finland
Mikko Kiviniemi Northern Savo Hospital District, Kuopio, Finland
Markku Voutilainen Hospital District of Southwest Finland, Turku, Finland
Mark Daly Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Ali Abbasi Abbvie, Chicago, IL, United States
Graham Heap Abbvie, Chicago, IL, United States
Jeffrey Waring Abbvie, Chicago, IL, United States
Nizar Smaoui Abbvie, Chicago, IL, United States
Fedik Rahimov Abbvie, Chicago, IL, United States
Anne Lehtonen Abbvie, Chicago, IL, United States
Keith Usiskin Celgene, Summit, NJ, United States/ Bristol Myers Squibb, New York, NY, United States
Tim Lu Genentech, San Francisco, CA, United States
Natalie Bowers Genentech, San Francisco, CA, United States
Danny Oh Genentech, San Francisco, CA, United States
Sarah Pendergrass Genentech, San Francisco, CA, United States
Linda McCarthy GlaxoSmithKline, Brentford, United Kingdom
Amy Hart Janssen Research & Development, LLC, Spring House, PA, United States
Meijian Guan Janssen Research & Development, LLC, Spring House, PA, United States
Jason Miller Merck, Kenilworth, NJ, United States
Rheumatology Group

Kari Eklund  Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Antti Palomäki  Hospital District of Southwest Finland, Turku, Finland
Pia Isomäki  Pirkanmaa Hospital District, Tampere, Finland
Laura Pirilä  Hospital District of Southwest Finland, Turku, Finland
Oili Kaipiainen-Seppänen  Northern Savo Hospital District, Kuopio, Finland
Johanna Huhtakangas  Northern Ostrobothnia Hospital District, Oulu, Finland
Nina Mars  Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland
Ali Abbasi  Abbvie, Chicago, IL, United States
Jeffrey Waring  Abbvie, Chicago, IL, United States
Fedik Rahimov  Abbvie, Chicago, IL, United States
Apinya Lertratanakul  Abbvie, Chicago, IL, United States
Nizar Smaoui  Abbvie, Chicago, IL, United States
Anne Lehtonen  Abbvie, Chicago, IL, United States
David Close  Astra Zeneca, Cambridge, United Kingdom
Marla Hochfeld  Celgene, Summit, NJ, United States/ Bristol Myers Squibb, New York, NY, United States
Natalie Bowers  Genentech, San Francisco, CA, United States
Sarah Pendergrass  Genentech, San Francisco, CA, United States
Jorge Esparza Gordillo  GlaxoSmithKline, Brentford, United Kingdom
Kirsi Auro  GlaxoSmithKline, Brentford, United Kingdom
Dawn Waterworth  Janssen Research & Development, LLC, Spring House, PA, United States
Fabiana Farias  Merck, Kenilworth, NJ, United States
Kirsi Kalpala  Pfizer, New York, NY, United States
Nan Bing  Pfizer, New York, NY, United States
Xinli Hu  Pfizer, New York, NY, United States
Pulmonology Group

Tarja Laitinen  
Pirkanmaa Hospital District, Tampere, Finland
Margit Pelkonen  
Northern Savo Hospital District, Kuopio, Finland
Paula Kauppi  
Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Hannu Kankaanranta  
Pirkanmaa Hospital District, Tampere, Finland
Terttu Harju  
Northern Ostrobothnia Hospital District, Oulu, Finland
Riitta Lahesmaa  
Hospital District of Southwest Finland, Turku, Finland
Nizar Smaoui  
Abbvie, Chicago, IL, United States
Alex Mackay  
Astra Zeneca, Cambridge, United Kingdom
Glenda Lassi  
Astra Zeneca, Cambridge, United Kingdom
Susan Eaton  
Biogen, Cambridge, MA, United States
Steven Greenberg  
Celgene, Summit, NJ, United States/ Bristol Myers Squibb, New York, NY, United States
Hubert Chen  
Genentech, San Francisco, CA, United States
Sarah Pendergrass  
Genentech, San Francisco, CA, United States
Natalie Bowers  
Genentech, San Francisco, CA, United States
Joanna Betts  
GlaxoSmithKline, Brentford, United Kingdom
Soumitra Ghosh  
GlaxoSmithKline, Brentford, United Kingdom
Kirsi Auro  
GlaxoSmithKline, Brentford, United Kingdom
Rajashree Mishra  
GlaxoSmithKline, Brentford, United Kingdom
Majd Mouded  
Novartis, Basel, Switzerland
Debby Ngo  
Novartis, Basel, Switzerland

Cardiometabolic Diseases Group

Teemu Niiranen  
The National Institute of Health and Welfare Helsinki, Finland
Felix Vaura  
The National Institute of Health and Welfare Helsinki, Finland
Veikko Salomaa  
The National Institute of Health and Welfare Helsinki, Finland
Kaj Metsärinne  
Hospital District of Southwest Finland, Turku, Finland
Jenni Aittokallio  
Hospital District of Southwest Finland, Turku, Finland
Mika Kähönen  
Pirkanmaa Hospital District, Tampere, Finland
Opthalmology Group

Kai Kaarniranta  Northern Savo Hospital District, Kuopio, Finland
Joni A Turunen  Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Terhi Ollila  Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Hannu Uusitalo  Pirkanmaa Hospital District, Tampere, Finland
Juha Karjalainen  Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Esa Pitkänen  Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Mengzhen Liu  Abbvie, Chicago, IL, United States
Heiko Runz  Biogen, Cambridge, MA, United States
Stephanie Loomis  Biogen, Cambridge, MA, United States
Erich Strauss  Genentech, San Francisco, CA, United States
Natalie Bowers  Genentech, San Francisco, CA, United States
Hao Chen  Genentech, San Francisco, CA, United States
Sarah Pendergrass  Genentech, San Francisco, CA, United States

Dermatology Group

Kaisa Tasanen  Northern Ostrobothnia Hospital District, Oulu, Finland
Laura Huilaja  Northern Ostrobothnia Hospital District, Oulu, Finland
Katarina Hannula-Jouppi  Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Teea Salmi  Pirkanmaa Hospital District, Tampere, Finland
Sirkku Peltonen  Hospital District of Southwest Finland, Turku, Finland
Leena Koulu  Hospital District of Southwest Finland, Turku, Finland
Nizar Smaoui  Abbvie, Chicago, IL, United States
Fedik Rahimov  Abbvie, Chicago, IL, United States
Anne Lehtonen  Abbvie, Chicago, IL, United States
David Choy  Genentech, San Francisco, CA, United States
Sarah Pendergrass  Genentech, San Francisco, CA, United States
Dawn Waterworth  Janssen Research & Development, LLC, Spring House, PA, United States
Kirsi Kalpala  Pfizer, New York, NY, United States
Sarah Pendergrass  Genentech, San Francisco, CA, United States
Karen S King  GlaxoSmithKline, Brentford, United Kingdom
Padhraig Gormley  GlaxoSmithKline, Brentford, United Kingdom
Jimmy Liu  GlaxoSmithKline, Brentford, United Kingdom
Karsten Sieber  Janssen Research & Development, LLC, Spring House, PA, United States
Amy Hart  Janssen Research & Development, LLC, Spring House, PA, United States
Meijian Guan  Janssen Research & Development, LLC, Spring House, PA, United States
Shicheng Guo  Janssen Research & Development, LLC, Spring House, PA, United States
Beryl Cummings  Maze Therapeutics, San Francisco, CA, United States
Matt Brauer  Maze Therapeutics, San Francisco, CA, United States
Jason Miller  Merck, Kenilworth, NJ, United States
Fabiana Farias  Merck, Kenilworth, NJ, United States
Jorge Del-Aguila  Merck, Kenilworth, NJ, United States
Kirill Shkura  Merck, Kenilworth, NJ, United States
Victor Neduva  Merck, Kenilworth, NJ, United States
Huilei Xu  Novartis, Basel, Switzerland
Amy Cole  Novartis, Basel, Switzerland
Jonathan Chung  Novartis, Basel, Switzerland
Jaison Jacob  Novartis, Basel, Switzerland
Katrina de Lange  Novartis, Basel, Switzerland
Jonas Zierer  Novartis, Basel, Switzerland
Xing Chen  Pfizer, New York, NY, United States
Åsa Hedman  Pfizer, New York, NY, United States
Clarence Wang  Sanofi, Paris, France
Ethan Xu  Sanofi, Paris, France
Franck Auge  Sanofi, Paris, France
Clement Chatelain  Sanofi, Paris, France
Deepak Rajpal  Sanofi, Paris, France
Dongyu Liu  Sanofi, Paris, France
Katherine Call  Sanofi, Paris, France
Nina Pitkänen  
Auria Biobank / University of Turku / Hospital District of Southwest Finland, Turku, Finland

**Biobank directors**

Lila Kallio  
Auria Biobank / University of Turku / Hospital District of Southwest Finland, Turku, Finland

Sirpa Soini  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Jukka Partanen  
Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland

Eero Punkka  
Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki

Raisa Serpi  
Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland

Kimmo Savinainen (acting) Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District, Tampere, Finland

Veli-Matti Kosma  
Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland

Teijo Kuopio  
Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland

**FinnGen Teams**

**Administration**

Anu Jalanko  
Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Huei-Yi Shen  
Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Risto Kajanne  
Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Mervi Aavikko  
Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

**Analysis**

Mitja Kurki  
Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland / Broad Institute, Cambridge, MA, United States
Juha Karjalainen Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Pietro Della Briotta Parolo Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Arto Lehiste Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Juha Mehtonen Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Wei Zhou Broad Institute, Cambridge, MA, United States
Masahiro Kanai Broad Institute, Cambridge, MA, United States
Mutaamba Maasha Broad Institute, Cambridge, MA, United States

Clinical Endpoint Development
Hannele Laivuori Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Aki Havulinna Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Susanna Lemmelä Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Tuomo Kiiskinen Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
L. Elisa Lahtela Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Communication
Mari Kaunisto Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

E-Science
Elina Kilpeläinen Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Timo P. Sipilä Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Oluwaseun Alexander Dada Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Awaisa Ghazal Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Anastasia Shcherban Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Genotyping
Kati Donner Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Timo P. Sipilä Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Sample Collection Coordination

Anu Loukola  
Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki

Sample Logistics

Päivi Laiho  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Tuuli Sistonen  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Essi Kaiharju  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Markku Laukkanen  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Elina Järvensivu  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Sini Lähteemäki  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Lotta Männikö  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Regis Wong  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Registry Data Operations

Hannele Mattsson  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Kati Kristiansson  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Susanna Lemmelä  
Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Sami Koskelainen  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Tero Hiekkalinna  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Teemu Paajanen  
THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Sequencing Informatics

Priit Palta  
Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Kalle Pärn  
Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Mart Kals  
Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Shuang Luo  
Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Vishal Sinha  
Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Trajectory
Tarja Laitinen  Pirkanmaa Hospital District, Tampere, Finland
Mary Pat Reeve  Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Marianna Niemi  University of Tampere, Tampere, Finland
Harri Siirtola  University of Tampere, Tampere, Finland
Javier Gracia-Tabuenca  University of Tampere, Tampere, Finland
Mika Helminen  University of Tampere, Tampere, Finland
Tiina Luukkaala  University of Tampere, Tampere, Finland
Iida Vähätalo  University of Tampere, Tampere, Finland

Data protection officer
Tero Jyrhämä  Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

FINBB – Finnish biobank cooperative
Marco Hautalahti
Johanna Mäkelä
Laura Mustaniemi
Mirkka Koivusalo
Sarah Smith
Tom Southerington

References:

Bots SH, SAE Peters and M Woodward. 2017. “Sex differences in coronary heart disease and


Sedlak, Tara, Romana Herscovici, Galen Cook-Wiens, Eileen Handberg, Janet Wei, Chrisandra Shufelt,


**Figure 1. Study design and workflow.**

**CHD biomarkers:** We identified CHD associated risk factors or biomarkers from UKB (Table 2 and Supplementary Table S1). Genome-wide association study (GWAS) was performed on UKB genotype data, separately in females and males, for 16 biomarkers. 

**Construction of CHDBioPRS:** We regressed CHD on the biomarkers in the UKB training data using elastic net Cox regression and retained 10 biomarkers. BioPRS is constructed by weighting each PRS by its coefficient in joint Cox regression model for CHD in UKB training data. CHDBioPRS is the sum of the standard CHDPRS and BioPRS where the weights are estimated from a Cox regression model predicting CHD within UKB validation data.
Figure 2. CHD prediction using PRSes in FinnGen and UKB Test samples. Both cohorts are disjoint from the samples used in construction of BioPRS and CHDBioPRS. Early onset = onset < 55 years of age.