Predicting the presence of coronary plaques featuring highrisk characteristics using polygenic risk scores and targeted proteomics in patients with suspected coronary artery disease

Møller, Peter Loof; Rohde, Palle Duun; Dahl, Jonathan Nørtoft ; Rasmussen, Laust Dupont; Nissen, Louise ; Schmidt, Samuel Emil; McGilligan, Victoria ; Gudbjartsson, Daniel F. ; Stefansson, Kari ; Holm, Hilma ; Bentzon, Jacob Fog ; Bøttcher, Morten ; Winther, Simon ; Nyegaard, Mette

Published in:
Genome Medicine

DOI (link to publication from Publisher):
10.1186/s13073-024-01313-8

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

Genome Medicine

RESEARCH

Open Access

Predicting the presence of coronary plaques featuring high-risk characteristics using polygenic risk scores and targeted proteomics in patients with suspected coronary artery disease

Peter Loof Møller[1,2], Palle Duun Rohde[2], Jonathan Nørtoft Dahl[3,4], Laust Dupont Rasmussen[3,10], Louise Nissen[3,4], Samuel Emil Schmidt[2], Victoria McGilligan[5], Daniel F. Gudbjartsson[6,7], Kari Stefansson[6,8], Hilma Holm[6], Jacob Fog Bentzon[4,9], Morten Bøttcher[3,4], Simon Winther[3,4] and Mette Nyegaard[2*]

Abstract

Background The presence of coronary plaques with high-risk characteristics is strongly associated with adverse cardiac events beyond the identification of coronary stenosis. Testing by coronary computed tomography angiography (CCTA) enables the identification of high-risk plaques (HRP). Referral for CCTA is presently based on pre-test probability estimates including clinical risk factors (CRFs); however, proteomics and/or genetic information could potentially improve patient selection for CCTA and, hence, identification of HRP. We aimed to (1) identify proteomic and genetic features associated with HRP presence and (2) investigate the effect of combining CRFs, proteomics, and genetics to predict HRP presence.

Methods Consecutive chest pain patients ($n = 1462$) undergoing CCTA to diagnose obstructive coronary artery disease (CAD) were included. Coronary plaques were assessed using a semi-automatic plaque analysis tool. Measurements of 368 circulating proteins were obtained with targeted Olink panels, and DNA genotyping was performed in all patients. Imputed genetic variants were used to compute a multi-trait multi-ancestry genome-wide polygenic score (GPS_{Mult}). HRP presence was defined as plaques with two or more high-risk characteristics (low attenuation, spotty calcification, positive remodeling, and napkin ring sign). Prediction of HRP presence was performed using the glmnet algorithm with repeated fivefold cross-validation, using CRFs, proteomics, and GPS_{Mult} as input features.

Results HRPs were detected in 165 (11%) patients, and 15 input features were associated with HRP presence. Prediction of HRP presence based on CRFs yielded a mean area under the receiver operating curve (AUC) ± standard error of 73.2 ± 0.1, versus 69.0 ± 0.1 for proteomics and 60.1 ± 0.1 for GPS_{Mult}. Combining CRFs with GPS_{Mult} increased prediction accuracy (AUC 74.8 ± 0.1 ($P = 0.004$)), while the inclusion of proteomics provided no significant improvement to either the CRF (AUC 73.2 ± 0.1, $P = 1.00$) or the CRF + GPS_{Mult} (AUC 74.6 ± 0.1, $P = 1.00$) models, respectively.

*Correspondence:
Mette Nyegaard
nyegaard@hst.aau.dk
Full list of author information is available at the end of the article

Conclusions In patients with suspected CAD, incorporating genetic data with either clinical or proteomic data improves the prediction of high-risk plaque presence.

Trial registration https://clinicaltrials.gov/ct2/show/NCT02264717 (September 2014).

Keywords High-risk coronary plaque, Coronary artery disease, Prediction, Genetics, Olink proteomics

Background

Coronary computed tomography angiography (CCTA) is guideline-endorsed for identifying patients with suspected obstructive coronary artery disease (CAD) [1, 2]. Additionally, CCTA enables risk stratification including detection of high-risk plaques (HRP) predisposing patients to adverse cardiac events [3–5]. HRPs are defined as having two or more characteristics on the CCTA associated with a higher likelihood of having an acute coronary syndrome (ACS) [6]. These characteristics include low attenuation plaque, spotty calcification, positive remodeling, and the napkin ring sign.

In patients with stable chest pain suggestive of obstructive CAD, pre-test probability (PTP) estimation is recommended to guide decisions about downstream testing [1, 2]. Classically, PTP estimation is based on sex, age, and chest pain typicality [1]. In general, both patient stratification, diagnostic value, and risk prediction have been improved by also considering clinical risk factors (CRFs) [7, 8] in the PTP estimation. The majority of de novo chest pain patients have low PTP and are recommended to undergo index testing by CCTA to identify potentially obstructive lesions [1, 2, 7]. Evidence of HRP characteristics has been documented in both patients with obstructive and non-obstructive lesions which CCTA outlines [4, 9], and as early treatment of patients with HRP characteristics could improve prognosis [10], an optimized clinical tool to guide CCTA utilization for HRP identification is warranted.

Based on large-scale proteomic panels and genotyping arrays yielding multi-protein models and genome-wide polygenic scores (GPSs), respectively, proteomics and genetics have been used for the prediction of obstructive CAD [11, 12], CAD-related traits [13], and general CAD risk management [14]. However, whether the combination of proteomic and genetic data improves the prediction accuracy of HRP presence is unknown. Further, a high PRS of CAD has previously been shown to be most predictive in patients younger than 55 years of age [15]. It is currently unknown whether this is also the case for HRP presence.

The primary aim of this study was to identify features (i.e., CRF, GPS, and proteins) associated with HRP presence and evaluate a combination of clinical, proteomic, and genetic data to predict the trait in de novo chest pain patients with suspected obstructive CAD. Secondly, we aimed to uncover any age-related variation in predictive features.

Methods

Study population

This is a sub-study of the Danish study of Non-Invasive testing in Coronary Artery Disease (Dan-NICAD) 1 study. The study protocol and main results have previously been described [16, 17]. In short, Dan-NICAD 1 was a prospective, multicenter cross-sectional study of 1675 patients with no previous history of CAD, low-intermediate PTP, and symptoms suggestive of obstructive CAD referred for initial testing by CCTA. On the day of CCTA, patients also underwent blood sampling, while symptoms and cardiac risk factors were registered. EDTA plasma was isolated from blood samples and stored at -80 °C.

CCTA add HRP definition

CCTA was performed on a 320-slice volume CT scanner (Aquilion One; Toshiba Medical Systems, Japan) following usual clinical guidelines. The presence and location of coronary plaques and CAD severity assessment were evaluated by a cardiologist [16]. CCTA plaque analysis was performed blinded to clinical risk factors, proteomics, and genetics, using the previously validated software Qangio CT (Research Edition ver. 3.1.4.2, Medis NL [18]).

Obstructive CAD was defined as a 50% diameter stenosis by visual assessment at CCTA. Plaques were evaluated by dedicated and trained core laboratory personnel (Cardiac Imaging Center, Department of Cardiology, Goedstrup Hospital, Denmark) for four qualitative high-risk characteristics: (1) positive remodeling (remodeling index > 1.10 calculated as vessel area at the location of plaque divided by reference vessel area in adjacent normal segments), (2) low-attenuation (non-calcified plaque with a plaque volume of > 1 mm [3] containing voxels with CT attenuation < 30 Hounsfield Units), (3) spotty calcification (< 3 mm calcium encased in non-calcified plaque), and (4) napkin ring sign (ring-like structure with lowest CT attenuation in the center of the plaque). The presence of HRP was defined as having coronary lesions with two or more high-risk

characteristics, as recommended by the CAD-RADS 2.0 system [19].

Clinical risk factor model

The clinical risk factor model was based on the features incorporated in the risk factor-weighted clinical likelihood (RF-CL) model of obstructive CAD [7]. The model is based on sex, age, angina typicality (typical, atypical, non-specific, and dyspnea), and the number of risk factors (family history of early CAD, smoking [never vs. active/former], dyslipidemia [receiving cholesterol-lowering medication], hypertension [receiving antihypertensive medication], and type 2 diabetes) ranging from 0 to 5. Models utilizing CRFs included all of the above features in their input.

Protein model

In total, 368 proteins were measured in EDTA plasma samples at Olink proteomics AB (Uppsala, Sweden) as previously described [20]. The protein analysis was done using four Olink Target 96 panels: the Cardiovascular II, Cardiovascular III, Inflammation, and Immune Response panel. The Cardiovascular II and III panels were chosen for their relevance to cardiovascular processes, while the Inflammation and Immune Response panels were included to broaden the search for associations. Samples marked in Olink's internal quality control steps were removed. Also, if the fraction of samples below the limit of detection (LOD) for a protein exceeded 20%, the protein was excluded from further analyses. For all included proteins, values were rank-normalized before adjusting for the collection box. A total of 300 proteins passed quality control and were included in the input to all models utilizing proteomic features.

Genetic model

Genotyping, quality control, and imputation have previously been described [20]. In short, genotyping used the Illumina Global Screening Array, followed by imputation using the Michigan Imputation Server. Single nucleotide polymorphisms (SNPs) with minor allele frequencies $\leq 1\%$ were excluded.

The genotype weights of the multi-trait multi-ancestry genome-wide polygenic score (GPS_{Mult}, https://www.pgscatalog.org/score/PGS003725/) by Patel et al. [21] were used to estimate genetic risk for CAD, as no dedicated HRP polygenic score exists. GPS_{Mult} aggregates multiple polygenic scores from CAD, BMI, ischemic stroke, diabetes mellitus, peripheral artery disease, glomerular filtration rate, systolic and diastolic blood pressure, LDL and HDL cholesterol, and triglycerides utilizing the 1,296,172 variants included in HapMap3.

Combined models

For combined models, the prediction was based on three groups of input features: (1) clinical risk factors (nine features), (2) proteins (300 features), and (3) GPS_{Mult} (one feature), with the name of each model reflecting the included feature groups.

Data analysis

The primary outcome was defined as the presence of HRP. For individual features (i.e., individual clinical risk factors, individual proteins, and GPS_{Mult}), performance was measured using the receiver operating characteristic (ROC) with the area under the curve (AUC) and 95% confidence interval (CI) implemented in the pROC [22] R package. Significance testing was performed using the Wilcoxon rank sum test and a Bonferroni corrected significance threshold of 0.05/315.

Models using more than one input feature were constructed using logistic regression with elastic net regularization [23, 24] implemented in the glmnet and caret [25] R packages, using weights to compensate for case–control imbalance. To limit the impact of random sampling noise, all models were based on 100 repeats of fivefold cross-validation (CV), foregoing a completely independent test set. ROC was used as a summary metric to select the optimal model. Hyperparameters were chosen from a combination of three alphas (0.1, 0.55, 1) and 10 lambdas ranging from 0.0001 to 1.

For all models, performance in every fold and repeat was stored, resulting in 500 AUC estimates for each model. Performance was reported as the mean ± standard error of all AUC estimates. As the AUC distributions for some models were not Gaussian, comparisons of model performance used the Kruskal–Wallis rank sum test to compare all models followed by Bonferroni corrected Dunn's test for individual post-hoc comparisons, reporting the adjusted *P* value.

Estimation of predictive performance in plaque subtypes was performed for all models. For each model, the HRP probability was averaged per individual across the 100 repeats, before calculation of the AUC with 95% CI for single features.

All analyses were performed within R [26] version 4.2.1. SHapley Additive explanation (SHAP) values were estimated using the fastshap [27] R package, using 100 Monte Carlo repetitions. Plotting was performed using ggplot2 [28].

Results

Study population

Of the 1675 eligible patients, 1462 (87%) had complete data on clinical risk factors, proteomics, genomics, and CCTA. Baseline demographics are shown in Table 1.

Table 1 Baseline information

	Overall ($n=1462$)		High-risk plaque ($n=165$)		No high-risk plaque ($n=1297$)	
Demographics						
Age, years	57 ± 9		60 ± 8		57 ± 9	
Males	699 (48%)		122 (74%)		577 (44%)	
Risk factors, n (%)						
Family history	527 (36%)		63 (38%)		464 (36%)	
Current/former smoker	778 (53%)		104 (63%)		674 (52%)	
Dyslipidemia	349 (24%)		42 (25%)		307 (24%)	
Hypertension	517 (35%)		69 (42%)		448 (35%)	
Type 2 diabetes	85 (6%)		16 (10%)		69 (5%)	
Type of chest pain, n (%)						
Typical angina	393 (27%)		63 (38%)		330 (25%)	
Atypical angina	497 (34%)		47 (28%)		450 (35%)	
Non-specific chest discomfort	271 (19%)		23 (14%)		248 (19%)	
Dyspnea	301 (21%)		32 (19%)		269 (21%)	
Laboratory tests						
Cholesterol medication	349 (24%)		42 (25%)		307 (24%)	
	Yes	No	Yes	No	Yes	No
Total cholesterol, mmol/L[a]	4.9 ± 1.2	5.5 ± 1.0	5.2 ± 1.6	5.8 ± 1.4	4.9 ± 1.1	5.5 ± 1.0
LDL cholesterol, mmol/L[a]	2.8 ± 1.1	3.4 ± 0.9	3.1 ± 1.2	3.7 ± 0.9	2.8 ± 1.0	3.4 ± 0.9
HDL cholesterol, mmol/L[a]	1.4 ± 0.4	1.5 ± 0.5	1.4 ± 0.4	1.4 ± 0.5	1.5 ± 0.4	1.5 ± 0.4
Triglyceride, mmol/L[a]	1.5 [1.0–2.1]	1.3 [0.9–1.9]	1.6 [1.0–2.3]	1.5 [1.1–2.3]	1.5 [1.0–2.1]	1.3 [0.9–1.9]
Measurements						
Blood pressure medication	517 (35%)		69 (42%)		448 (35%)	
	Yes	No	Yes	No	Yes	No
Systolic blood pressure, mm Hg[a]	143 ± 19	136 ± 18	144 ± 18	146 ± 19	143 ± 19	135 ± 18
Diastolic blood pressure, mm Hg[a]	84 ± 11	82 ± 11	83 ± 11	86 ± 11	85 ± 11	81 ± 11
Body mass index[a]	26.8 ± 4.2		26.3 ± 3.6		26.8 ± 4.3	
Obstructive CAD at CCTA	341 (23%)		134 (81%)		207 (16%)	
Coronary artery calcium score[a]	0 [0–81]		210 [52–624]		0 [0–42]	

Abbreviations: *CAD* coronary artery disease, *CCTA* coronary computed tomography angiography, *LDL* low-density lipoprotein, *HDL* high-density lipoprotein

Values are listed as mean ± standard deviation for normally distributed data; otherwise, the median and interquartile range are used

[a] Missing values were observed in total cholesterol $n=41$, LDL cholesterol $n=41$, HDL cholesterol $n=38$, triglyceride $n=46$, systolic and diastolic blood pressures $n=3$, body mass index $n=8$, coronary artery calcium score $n=1$

Baseline demographics of excluded patients are shown in Table S1, while the reason for exclusion is shown in Fig S1.

High-risk characteristics including positive remodeling, low-attenuation, spotty calcification, and napkin-ring sign were present in 309 (21%), 144 (10%), 181 (12%), and 36 (2%) patients, respectively (Fig. 1). In total, HRP presence was identified in 165/1462 (11%) patients, and 341/1462 patients (23%) had obstructive CAD. HRP presence was significantly correlated with obstructive CAD presence (Pearson's correlation = 0.49, $P<0.001$), with 134 patients having both.

Single clinical, protein, and genetic feature association with high-risk plaque

A total of 84 individual clinical, protein, and genetic features had a 95% CI lower limit of the AUC estimate above 50% (Fig. 2A). Among those, 14/84 (17%) features were significantly associated with HRP presence when accounting for multiple testing: three clinical risk factors (age, sex (male) and number of risk factors (0–5)), ten proteins (MMP12, TREM1, MMP-3, KIM1, CDCP1, PRSS8, LEP, GDF-15, PIgR, and COL1A1), and GPS_{Mult}. Stratifying patients by GPS_{Mult} quintiles revealed increased HRP prevalence with increasing polygenic

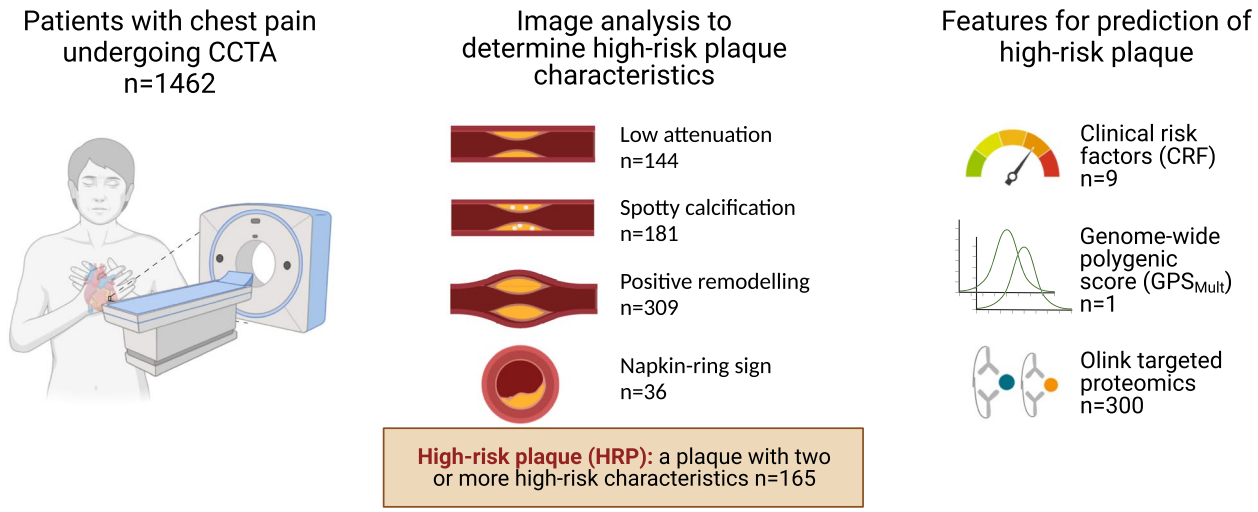

Fig. 1 Study design. 1462 patients underwent coronary computed tomography angiography (CCTA), followed by image analysis of high-risk plaque (HRP) characteristics. Finally, nine clinical risk factors, one multi-trait multi-ancestry genome-wide polygenic score (GPS_{Mult}), and 300 proteins were used to predict HRP presence

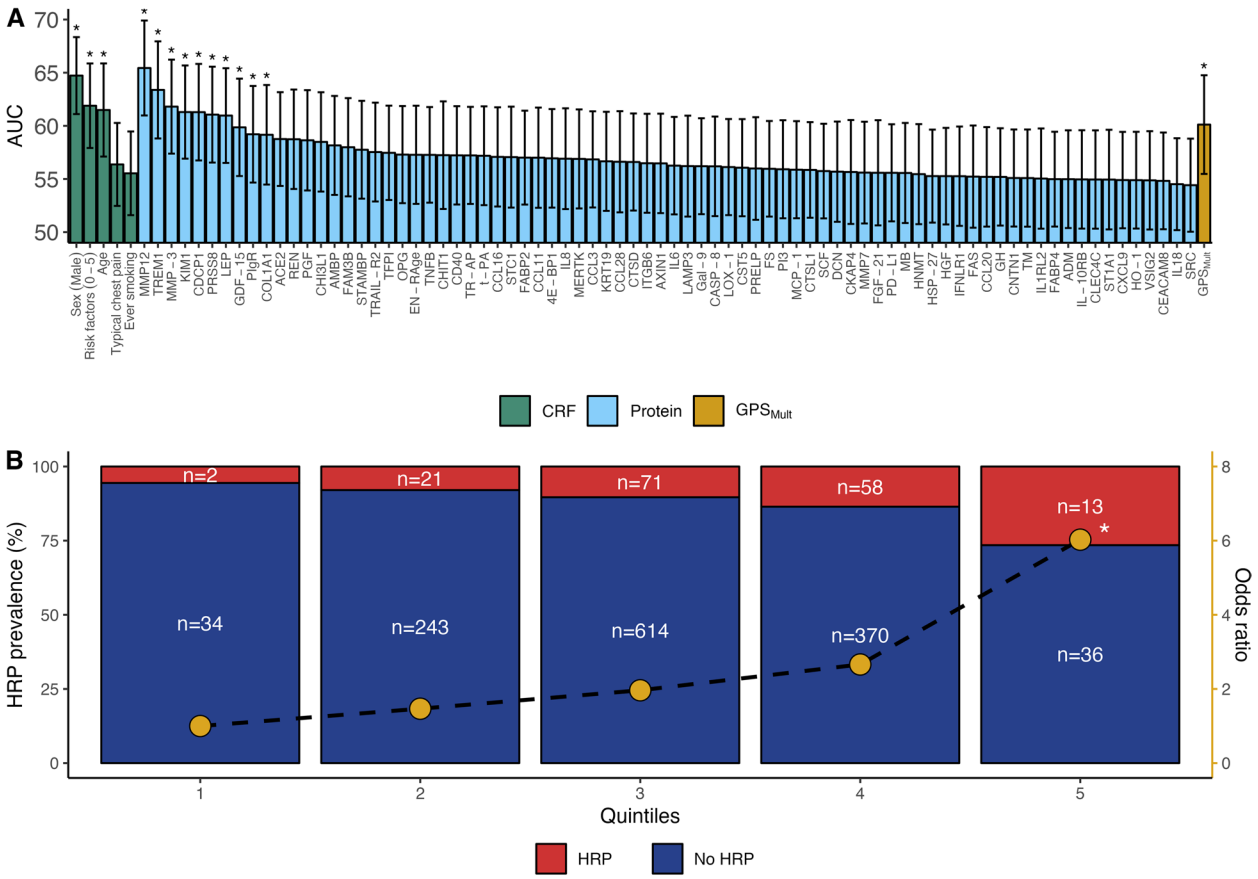

Fig. 2 Predictive performance of single features. **A** Area under the curve (AUC) for individual features, grouped by feature type. Error bars indicate a 95% confidence interval (CI). Asterisks indicate statistical significance after Bonferroni correction for multiple testing. CRF, clinical risk factors; GPS_{mult}, multi-trait multi-ancestry genome-wide polygenic score. Only single features with a lower limit of 95% CI above 50% are shown. **B** Prevalence of high-risk plaque (HRP) stratified by GPS_{Mult} quintiles. Odds ratios are calculated using the first quintile as a reference. Asterisk indicates a statistically significant difference in the estimated odds ratio compared to the reference quintile

score, with the highest quintile having an odds ratio of 6.03 ($P=0.02$) relative to the lowest quintile (Fig. 2B).

Combined prediction models of high-risk plaque

Compared to the model based on CRF features, the prediction accuracy of HRP presence by the protein model was lower (AUC 73.2 ± 0.1 vs. 69.0 ± 0.1, $P<0.001$) (Fig. 3A). The GPS_{Mult} model had an AUC of 60.1 ± 0.1, which was inferior to both the CRF and protein models.

Combining GPS_{Mult} with CRFs or protein features increased prediction accuracy of HRP presence (CRF + GPS_{Mult}: 74.8 ± 0.1, $P=0.004$; Protein + GPS_{Mult} = 71.0 ± 0.1, $P<0.001$). Combining CRF and protein features leads to the inclusion of eight features in total, four of these being proteins, but no improvement in the accuracy of prediction was observed compared to the CRF model (73.2 ± 0.1, $P=1.00$). Similarly, a full model including all feature groups did not improve the CRF + GPS_{Mult} model (74.6 ± 0.1, $P=1.00$).

Across all models, using each model for predicting individual high-risk plaque characteristics showed the lowest predictive performance for low attenuation plaques and the highest for napkin ring sign (Table 2).

Supplementary analysis defining HRP as 1- or 3-feature positive plaques (instead of 2-feature) found CRF to have the best predictive ability in 3-feature HRP ($n=29$), while CRF + GPS_{Mult} had the best predictive ability for 1-feature HRP ($n=388$, Table S2).

Age stratification

Stratifying the cohort into patients above and below 55 years of age showed an HRP prevalence of 8.0% (50/627) in patients ≤ 55 years and 13.8% (115/835) in patients > 55 years. For all non-protein models, this stratification improved HRP prediction in patients ≤ 55 years (Fig. 3) relative to patients > 55 years. Especially the GPS_{Mult} among patients ≤ 55 displayed improved prediction accuracy significantly ($AUC_{\leq 55}=66.5 \pm 0.4$ vs. AUC

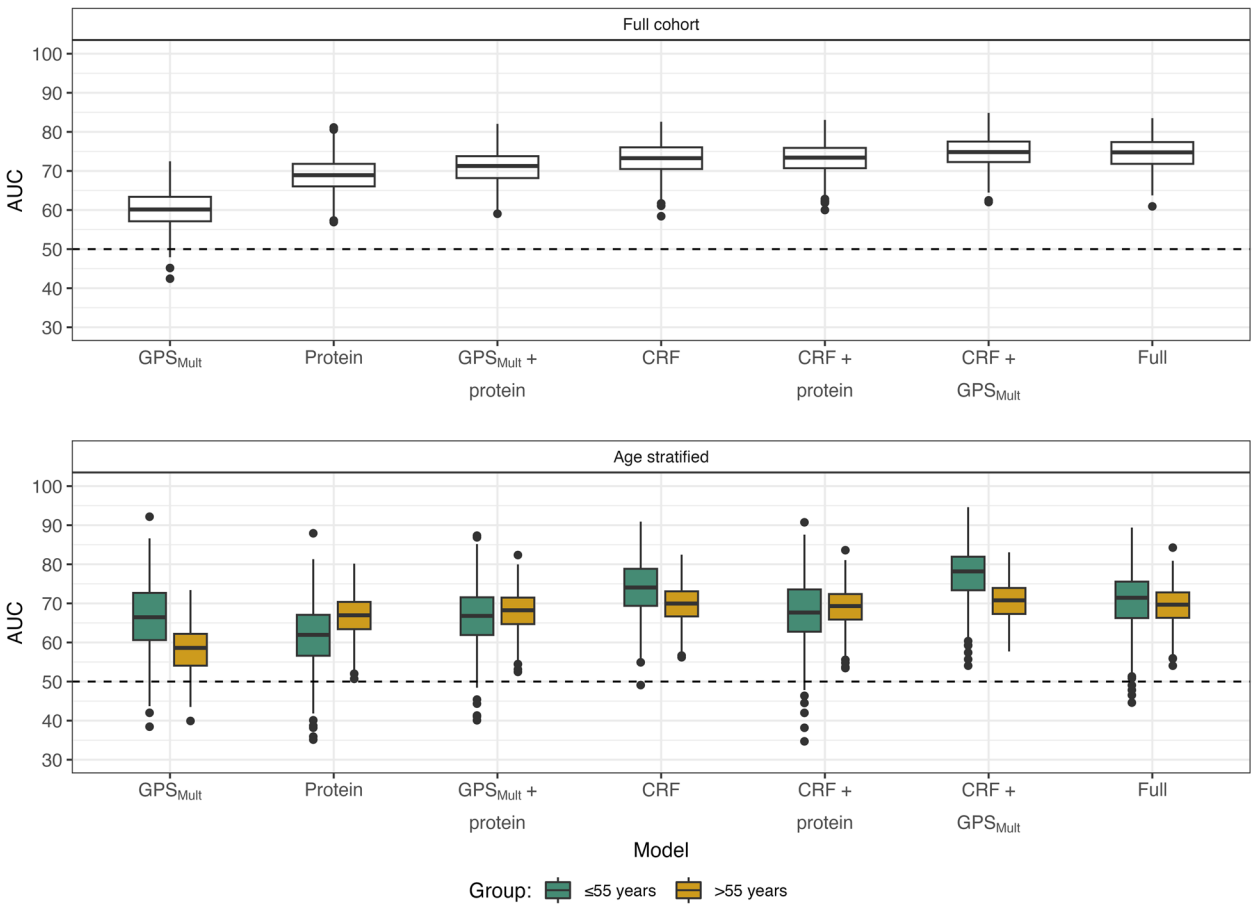

Fig. 3 Predictive performance of individual and combined models. Models included features from up to three feature groups, as shown on the *x*-axis, revealing the CRF + GPS_{Mult} model to be the most predictive in the full cohort. Stratifying patients by age group resulted in improved GPS_{Mult} performance in the group ≤ 55 years of age, while the CRF + GPS_{Mult} models were best in both groups. AUC, area under the curve; CRF, clinical risk factor; GPS_{Mult}, multi-trait multi-ancestry genome-wide polygenic score; Full, CRF + protein + GPS_{Mult}

Table 2 Plaque subtype prediction

Models	Low attenuation ($n=144$)	Spotty calcification ($n=181$)	Positive remodeling ($n=309$)	Napkin ring sign ($n=36$)
CRF	71.1 [66.8–75.4]	72.7 [69.0–76.3]	71.9 [68.7–75.0]	76.9 [69.4–84.3]
Protein	67.4 [62.9–71.8]	69.3 [65.3–73.4]	69.0 [65.6–72.3]	74.2 [66.5–81.9]
GPS_{Mult}	59.2 [54.3–64.1]	60.2 [55.9–64.5]	62.7 [59.2–66.1]	61.1 [51.6–70.6]
CRF + GPS_{Mult}	72.8 [68.6–77.0]	75.1 [71.7–78.5]	74.9 [71.9–77.8]	78.0 [70.0–85.9]
Protein + GPS_{Mult}	69.1 [64.9–73.3]	71.5 [67.7–75.3]	71.9 [68.8–75.0]	76.3 [69.3–83.3]
CRF + Protein	71.1 [66.6–75.6]	73.0 [69.3–76.8]	71.5 [68.3–74.7]	79.3 [72.7–85.9]
Full	72.3 [68.3–76.4]	75.0 [71.5–78.6]	74.9 [72.0–77.9]	79.5 [72.8–86.2]

CRF clinical risk factor, GPS_{Mult} multi-trait multi-ancestry genome-wide polygenic score

Numbers are the area under the curve with a 95% confidence interval

$_{>55}$: 58.3 ± 0.2, $P<0.001$). The best-performing model in both age groups was based on CRF and GPS_{Mult}, yielding improved prediction in younger patients (AUC $_{\leq 55}$ = 77.3 ± 0.3 vs. $AUC_{>55}$ = 70.7 ± 0.02 ($P<0.001$)).

Predicting individual high-risk plaque characteristics in patients ≤ 55 years of age showed similar results as in the full cohort (Table S3).

Comparing the impact of individual features (utilizing SHAP values) on model output in the CRF + GPS_{Mult} models across the three groups (full cohort, ≤ 55 years of age, and > 55 years of age), sex was consistently the most influential single feature on HRP presence (Fig. 4A). Comparing the impact of CFRs to GPS_{Mult} revealed GPS_{Mult} to be equally important to the sum of risk factors. For individual risk factors, family history of early CAD appeared protective in the full cohort, smoking had an impact in patients > 55 years of age, and cholesterol medication was protective in patients > 55 years of age and the full cohort (Fig. 4A). Finally, the use of antihypertensive medication and having diabetes both had a low impact on all model predictions. Typical chest pain symptoms indicated increased HRP risk in all groups, while nonspecific chest pain in patients ≤ 55 years of age and other chest pain in patients > 55 years of age and the full cohort indicated reduced HRP risk.

Examining the individual SHAP values (Fig. 4B) revealed multiple features (age, number of risk factors, and GPS_{Mult}) with the potential to be more impactful than sex for a single patient although they were inferior from a cohort perspective.

Discussion

The main finding of the study was that HRP presence can be predicted using clinical risk factors, whereas genetic data primarily serves as a strong source of complementary information, resulting in improved HRP prediction in the complete cohort and especially among patients below 55 years of age the prediction accuracy was significantly improved. Finally, we did not find that our proteomic data improved the overall discrimination, despite having better individual prediction accuracy than genetic data.

Looking at individual input features, 14 features were significantly associated with HRP presence, with well-known clinical risk factors and proteins showing superior discrimination. Our findings are not surprising as several proteins are correlated with sex and age which independently correlates to HRP presence (Fig. 2A). In our study, we used protein measurements without adjusting the protein level for age and sex, as we wanted to retain these signals for models which did not incorporate age and sex directly.

Regarding the combined models, CRF features were the most predictive feature group for HRP presence. In contrast, Bom et al. [13] reported 196 patients with 22% HRP prevalence and utilized four Olink panels, three of which were also analyzed in our study, and found proteins to be predictive of HRP presence (AUC of 79 ± 1) with impaired prediction accuracy by CRFs alone (AUC of 65 ± 4). The difference in CRF predictive performance between our studies could be explained by our study having more detailed CRF data, e.g., the number of risk factors as implemented in the RF-CL model. Furthermore, the higher predictive performance of proteomics observed by Bom et al. may be explained by their inclusion of the targeted Cardiometabolic Olink panel which was not included in our study. This highlights the possibility of future discoveries as proteomic panels include an increasing number of proteins.

Due to the rarity of cohorts with detailed high-risk plaque characteristics, no suitable GWAS summary statistic exists for the calculation of a traditional polygenic score for HRP. Instead, we leveraged GPS_{Mult}, an existing polygenic score for CAD, which incorporates genetic information about multiple known CAD risk factors and from multiple ancestries. GPS_{Mult} showed the greatest

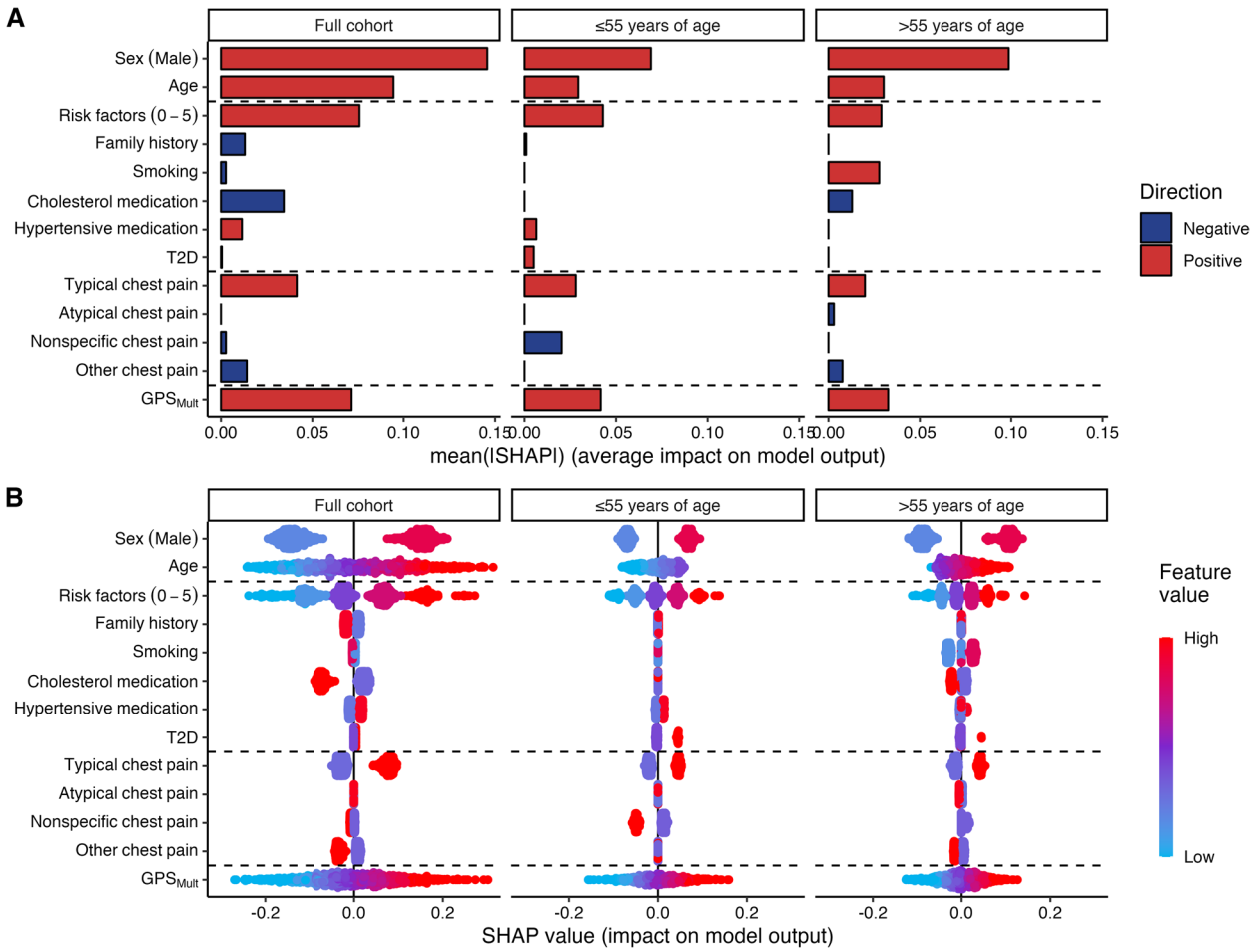

Fig. 4 Feature importance of the CRF + GPS_{Mult} models across cohort groups. **A** Mean absolute SHAP values, representing the average importance of input features on model prediction, with blue features leading to lower risk and red features leading to higher risk. **B** Patient-level SHAP values, showing how impactful some features can be in extreme cases. SHAP, Shapley additive explanation; T2D, type 2 diabetes mellitus; GPS_{Mult}, multi-trait multi-ancestry genome-wide polygenic score

prediction accuracy in patients below 55 years of age. This is consistent with previous findings by Mars et al. reporting PRS for CAD prediction to have lower performance in people above 55 years of age [15] and also explains the somewhat low predictive ability of GPS_{Mult} by itself in our study (mean age of 60 years).

Interestingly, we found that GPS_{Mult} was able to improve prediction when combined with either CRFs or proteins, indicating that genetics carry additional information about existing or novel aspects of HRP risk. Meanwhile, the inclusion of proteins with CRFs did not improve prediction, suggesting that the proteins assessed in our study do not contain additional information about HRP risk. It is possible that future improvements in HRP prediction could arise from increasing the number of included plasma proteins, using, e.g., SomaScan 7K or Olink Explore HT.

One of the strengths of this study is the consecutive enrollment of nearly 1500 patients with detailed CCTA readings allowing for the discovery of HRP presence and specific plaque characteristics. Additionally, the Danish healthcare system is likely to reduce referral bias, as no direct payment is needed for citizens to undergo diagnostic testing and therefore our findings represent an all-comer population.

The primary limitation of the study is the lack of external validation. Instead, this study utilized internal testing through repeated cross-validation, enabling visualization of the uncertainty of our performance estimates (Fig. 3). Additionally, this study investigated only 300 proteins, meaning that additional studies, using larger, more explorative panels, are required to investigate the remaining plasma proteome.

Conclusion

Genetic data can be used to improve both clinical risk factors and proteomic models for the prediction of HRP presence, especially in patients below 55 years of age. However, a model combining both clinical risk factors and proteomics did not improve high-risk plaque identification, despite both types of data being predictive on their own.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-024-01313-8.

Additional file 1: Table S1. Baseline information. **Table S2.** Analysis of 1, 2, and 3 feature plaques. **Table S3.** Plaque subtype prediction in patients ≤55 years of age. **Fig S1.** Patient inclusion.

Additional file 2: Table S4. Summary level data of number of risk factors, proteomics and GPS_{Mult} stratified by case/control status.

Acknowledgements
The authors would like to thank the Dan-NICAD study collaborators, including study nurses and clinical staff at the enrollment centers.

Authors' contributions
The authors confirm the following contributions to the paper: study conception and design: PLM, PDR, MN, MB, SW; data collection: LN, MB, SW, VM, DFG, KS, HH; analysis and interpretation of results: JND, LDR, PLM, PDR, MN; draft manuscript preparation: PLM, PDR, MN, JFB, SES. All authors reviewed the results and read and approved the final manuscript.

Funding
VM was supported by funding from the European Union Regional Development Fund (ERDF) EU Sustainable Competitiveness Programme for Northern Ireland; Northern Ireland Public Health Agency (HSC R&D). SW was supported by the Novo Nordisk Foundation Clinical Emerging Investigator grant (NNF21OC0066981). MN was supported by the Novo Nordisk Foundation Start Package grants for faculty recruitment (NNF0071050).

Availability of data and materials
Summary level data of number of risk factors, proteomics, and GPS_{Mult} are available in Table S4. Individual level data cannot be made publicly available.

Declarations

Ethics approval and consent to participate
The study follows the principles outlined in the Declaration of Helsinki. The study was approved by the Central Denmark Regional Committee on Health Research Ethics (record number: 1–10-72–190-14) and the Danish Data Protection Agency (record number: 1–16-02–345-14). Informed written consent was obtained from all patients prior to enrollment. The trial was registered at https://www.clinicaltrials.gov (unique identifier: NCT02264717).

Consent for publication
Not applicable.

Competing interests
MB discloses advisory board participation for NOVO Nordisk, AstraZeneca, Pfizer, Boehringer Ingelheim, Bayer, Sanofi, Novartis, AMGEN, CLS-Behring, and Acarix. SES is a part-time consultant in Acarix and minor shareholder of Acarix. DFG, KS, and HH are employees of deCODE genetics/Amgen Inc. The remaining authors have no competing interests.

Author details
[1]Department of Biomedicine, Aarhus University, Aarhus, Denmark. [2]Department of Health Science and Technology, Aalborg University, Aalborg, Denmark. [3]Department of Cardiology, Gødstrup Hospital, Herning, Denmark. [4]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark. [5]Personalized Medicine Centre, School of Medicine, Ulster University, Derry, Northern Ireland. [6]deCODE Genetics/Amgen, Inc, Reykjavik, Iceland. [7]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [8]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [9]Centro Nacional de Investigaciones Cardiovasculares, Madrid, Spain. [10]Department of Cardiology, Aalborg University Hospital, Aalborg, Denmark.

Received: 29 August 2023 Accepted: 12 March 2024
Published online: 20 March 2024

References

1. Neumann FJ, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes: the task force for the diagnosis and management of chronic coronary syndromes of the European Society of Cardiology (ESC). Eur Heart J. 2020;41:407–77.
2. Gulati M, et al. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR guideline for the evaluation and diagnosis of chest pain: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. J Am Coll Cardiol. 2021;144:e368–454.
3. Shaw LJ, et al. Society of cardiovascular computed tomography / North American Society of Cardiovascular imaging - expert consensus document on coronary CT imaging of atherosclerotic plaque. J Cardiovasc Comput Tomogr. 2021;15:93–109.
4. Williams MC, et al. Coronary artery plaque characteristics associated with adverse outcomes in the SCOT-HEART study. J Am Coll Cardiol. 2019;73:291–301.
5. Ferencik M, et al. Coronary atherosclerosis, cardiac troponin, and interleukin-6 in patients with chest pain: the PROMISE trial results. JACC Cardiovasc Imaging. 2022;15:1427–38.
6. Motoyama S, et al. Plaque characterization by coronary computed tomography angiography and the likelihood of acute coronary events in mid-term follow-up. J Am Coll Cardiol. 2015;66:337–46.
7. Winther S, et al. Incorporating coronary calcification into pre-test assessment of the likelihood of coronary artery disease. J Am Coll Cardiol. 2020;76:2421–32.
8. Winther S, et al. Coronary calcium scoring improves risk prediction in patients with suspected obstructive coronary artery disease. J Am Coll Cardiol. 2022;80:1965–77.
9. Senoner T, et al. Added value of high-risk plaque criteria by coronary CTA for prediction of long-term outcomes. Atherosclerosis. 2020;300:26–33.
10. Dawson LP, Lum M, Nerleker N, Nicholls SJ, Layland J. Coronary atherosclerotic plaque regression: JACC state-of-the-art review. J Am Coll Cardiol. 2022;79:66–82.
11. Klarin D, Natarajan P. Clinical utility of polygenic risk scores for coronary artery disease. Nat Rev Cardiol. 2021;19(5):291–301.
12. Ferrannini E, et al. Differential proteomics of cardiovascular risk and coronary artery disease in humans. Front Cardiovasc Med. 2022;8:790289.
13. Bom MJ, et al. Predictive value of targeted proteomics for coronary plaque morphology in patients with suspected coronary artery disease. EBioMedicine. 2019;39:109–17.
14. Roberts R, Campillo A, Schmitt M. Prediction and management of CAD risk based on genetic stratification. Trends Cardiovasc Med. 2020;30:328–34.
15. Mars N, 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;26:549–57.
16. Nissen L, et al. Danish study of Non-Invasive testing in Coronary Artery Disease (Dan-NICAD): study protocol for a randomised controlled trial. Trials. 2016;17:262.
17. Nissen L, et al. Diagnosing coronary artery disease after a positive coronary computed tomography angiography: the Dan-NICAD open label, parallel, head to head, randomized controlled diagnostic accuracy trial of cardiovascular magnetic resonance and myocardial perfusion scintigraphy. Eur Heart J Cardiovasc Imaging. 2018;19:369–77.
18. De Graaf MA, et al. Automatic quantification and characterization of coronary atherosclerosis with computed tomography coronary angiography:

cross-correlation with intravascular ultrasound virtual histology. Int J Cardiovasc Imaging. 2013;29:1177–90.
19. Cury RC, et al. CAD-RADS^{TM} 2.0 - 2022 Coronary artery disease-reporting and data system: an expert consensus document of the Society of Cardiovascular Computed Tomography (SCCT), the American College of Cardiology (ACC), the American College of Radiology (ACR), and the North America Society of Cardiovascular Imaging (NASCI). J Cardiovasc Comput Tomogr. 2022;16:536–57.
20. Møller PL, et al. Sortilin as a biomarker for cardiovascular disease revisited. Front Cardiovasc Med. 2021;8:652584.
21. Patel AP, et al. A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. Nat Med. 2023;29(7):1793–803.
22. Robin X, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
23. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33:1–22.
24. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. J Stat Softw. 2011;39:1–13.
25. Kuhn M. Building Predictive Models in R Using the caret Package. J Stat Softw. 2008;28(5):1–26.
26. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Version 4.2.1. 2023. https://www.R-project.org/.
27. Greenwell B. fastshap: Fast Approximate Shapley Values. R package version 0.1.0. 2023. https://CRAN.R-project.org/package=fastshap.
28. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.

Publisher's Note