Circulation: Genomic and Precision Medicine

ORIGINAL ARTICLE

Epigenetic Contributions to Clinical Risk Prediction of Cardiovascular Disease

Aleksandra D. Chybowska, MSci; Danni A. Gadd, MSc; Yipeng Cheng, MSc; Elena Bernabeu, PhD; Archie Campbell, MA; Rosie M. Walker, PhD; Andrew M. McIntosh, PhD; Nicola Wrobel, BSc; Lee Murphy, MSc; Paul Welsh, PhD; Naveed Sattar, PhD; Jackie F. Price, MB, ChB, MD; Daniel L. McCartney, PhD; Kathryn L. Evans, PhD; Riccardo E. Marioni, PhD

BACKGROUND: Cardiovascular disease (CVD) is among the leading causes of death worldwide. The discovery of new omics biomarkers could help to improve risk stratification algorithms and expand our understanding of molecular pathways contributing to the disease. Here, ASSIGN—a cardiovascular risk prediction tool recommended for use in Scotland—was examined in tandem with epigenetic and proteomic features in risk prediction models in ≥12 657 participants from the Generation Scotland cohort.

METHODS: Previously generated DNA methylation–derived epigenetic scores (EpiScores) for 109 protein levels were considered, in addition to both measured levels and an EpiScore for cTnI (cardiac troponin I). The associations between individual protein EpiScores and the CVD risk were examined using Cox regression (n_{cases}≥1274; $n_{controls}$≥11 383) and visualized in a tailored R application. Splitting the cohort into independent training (n=6880) and test (n=3659) subsets, a composite CVD EpiScore was then developed.

RESULTS: Sixty-five protein EpiScores were associated with incident CVD independently of ASSIGN and the measured concentration of cTnI (P<0.05), over a follow-up of up to 16 years of electronic health record linkage. The most significant EpiScores were for proteins involved in metabolic, immune response, and tissue development/regeneration pathways. A composite CVD EpiScore (based on 45 protein EpiScores) was a significant predictor of CVD risk independent of ASSIGN and the concentration of cTnI (hazard ratio, 1.32; P=3.7×10^{-3}; 0.3% increase in C-statistic).

CONCLUSIONS: EpiScores for circulating protein levels are associated with CVD risk independent of traditional risk factors and may increase our understanding of the etiology of the disease.

Key Words: biomarkers ■ cardiovascular diseases ■ epigenomics ■ multiomics ■ troponin

See Editorial by Bozack et al

For the past 20 years, cardiovascular disease (CVD) has been among the leading causes of mortality and morbidity worldwide. Given that many CVD cases are preventable, it is important to identify at-risk individuals early, when an intervention is most likely to be effective, and translate this knowledge into preventative strategies.[1,2] Although there are many CVD risk prediction algorithms, currently, they have limited predictive performance.[3] It may be possible to improve on that by discovering novel factors strongly associated with the disease, for example, the type and the concentrations of proteins expressed as a response to the damage to the cardiovascular system.

Correspondence to: Riccardo E. Marioni, PhD, Centre for Genomic and Experimental Medicine, The University of Edinburgh, Western General Hospital, Crewe Rd, Edinburgh EH4 2XU, United Kingdom. Email riccardo.marioni@ed.ac.uk

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/CIRCGEN.123.004265.
For Sources of Funding and Disclosures, see page 45.

Circulation: Genomic and Precision Medicine is available at www.ahajournals.org/journal/circgen

Nonstandard Abbreviations and Acronyms

ADM	adrenomedullin
cTnC	cardiac troponin C
cTnI	cardiac troponin I
cTnT	cardiac troponin T
CVD	cardiovascular disease
DNAm	DNA methylation
EpiScores	epigenetic scores
GDF15	growth differentiation factor 15
GS	Generation Scotland
HR	hazard ratio
NT-proBNP	N-terminal pro-B-type natriuretic peptide
PH	proportional hazard

Several proteins have been highlighted as possible biomarkers for CVD. These include GDF15 (growth differentiation factor 15), NT-proBNP (N-terminal pro-B-type natriuretic peptide), and ADM (adrenomedullin).[4–7] An established and highly sensitive marker of myocardial damage is cardiac troponin.[8] It is a complex of 3 proteins, namely, cTnI (cardiac troponin I), cTnT (cardiac troponin T), and cTnC (cardiac troponin C) regulating the contraction of the cardiac muscle. Cardiac forms of troponin T[9,10] and troponin I are expressed almost exclusively in the heart.[11] Following myocyte damage, cardiac troponin enters the circulation and can be detected in blood samples. A high-sensitivity cardiac troponin test plays a role in the rapid diagnosis of myocardial infarction.[8] Low-grade elevations in cardiac troponin are associated with an increased risk of CVD.[8]

Individual differences in protein concentration can be well captured by DNA methylation (DNAm). DNAm is a type of epigenetic modification characterized by the addition of methyl groups to DNA. Typically, the methyl group is added to cytosine-phosphate-guanine dinucleotides that are found mostly (but not exclusively) in gene promoters.[12] Blocking promoters, to which activating transcription factors should bind to initiate transcription, is one of the mechanisms by which DNAm can precisely regulate gene expression.[13] Conversely, changes in DNAm patterns can also be a result of changes in gene expression and chromatin state.[14,15]

DNAm-based proxies for protein levels are referred to as protein epigenetic scores (EpiScores) and are broadly analogous to polygenic risk scores. These methylation scores can be derived from penalized linear regression models of protein concentrations. Due to their temporal stability, protein EpiScores may exhibit stronger associations with disease outcomes than singular protein measurements, which are known to fluctuate between measurements.[16–19] We have shown that EpiScores for 109 circulating protein levels are associated with the time to diagnosis for a host of leading causes of morbidity and mortality, including cardiovascular outcomes.[20] Protein EpiScores are, therefore, useful biomarker tools for disease risk stratification.

Here, we examine whether protein EpiScores, calculated for ≥12 657 participants of the Generation Scotland (GS), study can augment predictions made by a CVD risk calculator developed for use in Scotland (ASSIGN[21]). We first run individual Cox proportional hazard (PH) models to discover relationships between individual protein EpiScores and incident CVD. We then create a CVD EpiScore (based on the protein EpiScores) and test the additional predictive performance offered by it for CVD risk stratification. A graphical overview of the analyses is presented in Figure 1.

METHODS

All methods are described in the Supplemental Material. A key resource in this study, GS, is a family-based research initiative focusing on genetic and environmental factors influencing health. Briefly, from 2006 to 2011, eligible individuals were selected from participating general medical practices in Scotland and invited at random to take part in the study.[22] All participants provided written informed consent for research. The study received ethical approval from the National Health Service Tayside Committee on Medical Research Ethics (REC reference number: 05/S1401/89). The GS data set is not publicly available as it contains information that could compromise participant consent and confidentiality. However, the data, research materials, and analytical methods will be made accessible to other researchers for the purpose of replicating the findings. Access will be granted upon successful project application to the GS Access Committee and obtaining ethical approval for accessing linked health data from NHS Scotland. Instructions for accessing GS data can be found at https://www.ed.ac.uk/generation-scotland/for-researchers/access; the GS Access Request Form can be downloaded from this site.

RESULTS

Clinical Risk Prediction Tools

ASSIGN scores were calculated for 16 366 individuals with nonmissing risk factor data. To meet the PH assumption of the Cox model, the data set was filtered to individuals aged between 30 and 70 years (results split by decade are presented in Table SI) and trimmed of outliers (points beyond 3 SDs of the mean; n=181). This left a cohort of 12 790 individuals, which was further filtered to records with nonmissing concentrations of cTnI (n=12 657). Table 1 summarizes the training, test, and full data sets.

Incremental Model Using Cardiac Troponin and Cardiac Troponin EpiScores

We tested whether concentrations of cardiac troponin were associated with CVD risk above ASSIGN over 16 years of follow-up. While the measured concentration of cTnI was associated with a hazard ratio (HR) of 1.20 per SD increase in the full (n=12 657) cohort (95% CI,

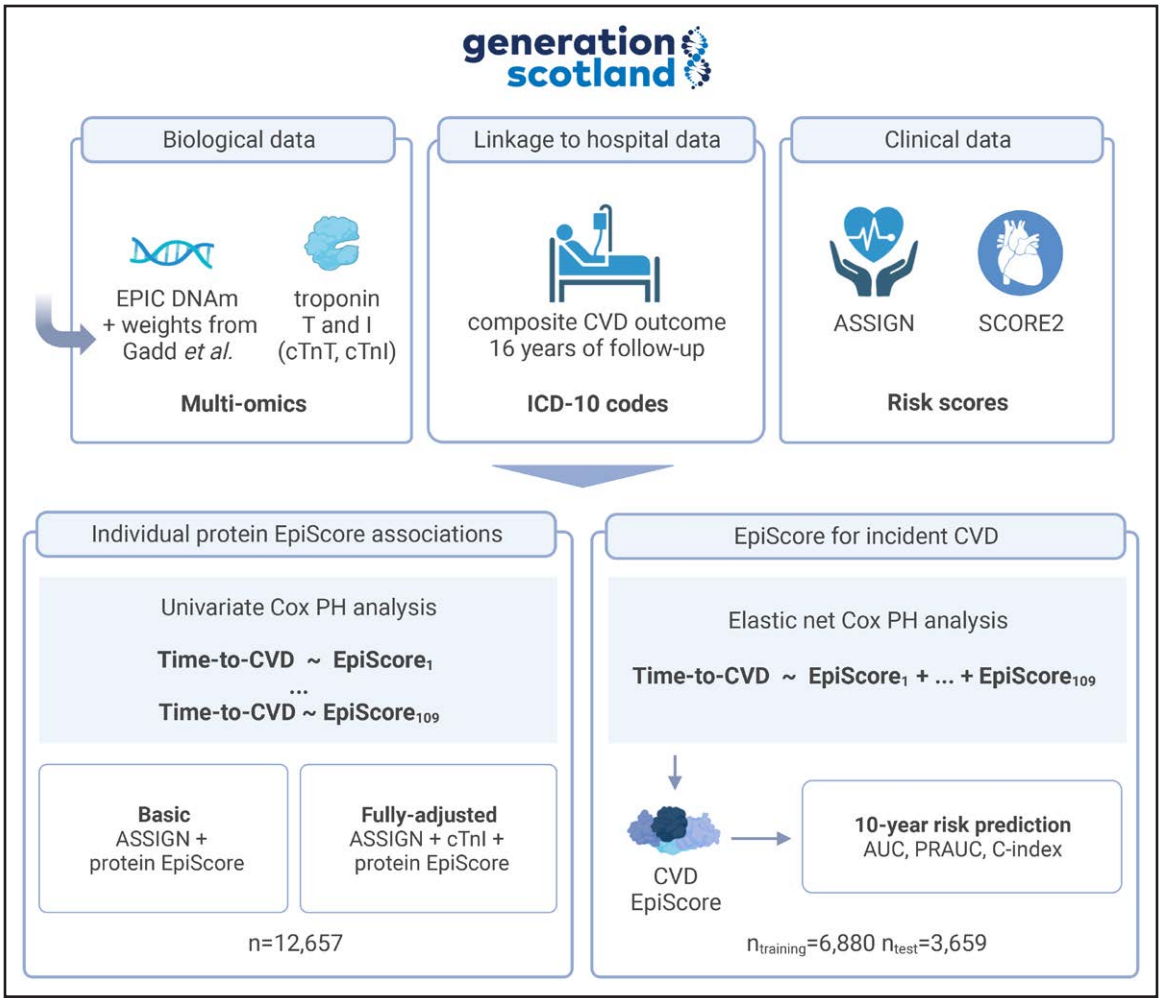

Figure 1. Project overview.
A series of Cox proportional hazard (PH) models were run to model the relationship between time-to-cardiovascular disease (CVD) and 109 protein epigenetic scores (EpiScores). Basic models were adjusted for the ASSIGN score, whereas fully adjusted models also included the concentration of cTnI (cardiac troponin I). This was followed by a prediction analysis where a composite protein EpiScore was trained. The CVD EpiScore was derived using elastic net and 109 protein EpiScores as possible input features. The score was assessed in the test sample to quantify the additional predictive performance offered by it over and above ASSIGN and SCORE2. The test Cox PH models were adjusted for age, sex, cTnI, and the CVD EpiScore, with time-to-CVD as the outcome. ASSIGN indicates the cardiovascular risk score chosen for use by SIGN (Scottish Intercollegiate Guidelines Network) and Scottish Government Health Directorates; AUC, area under the receiver operating characteristic curve; cTnT, cardiac troponin T; PRAUC, area under the precision recall curve; and SCORE2, an algorithm derived, calibrated, and validated to predict 10-year risk of first-onset CVD in European populations. Created with BioRender.com

1.13–1.29; P=1.9×10^{-8}), an EpiScore generated for cTnI (see Methods for details) was not associated with the measured concentrations in the n=3659 test set (incremental R^2, 0.027%; P=0.31) and did not predict CVD risk in Cox models adjusted for ASSIGN in the same test set (P=0.59). For that reason, it was not considered a feature in the generation of the composite CVD score.

Incremental Model Using EpiScores for Plasma Protein Levels

We then tested whether 109 protein EpiScores generated by Gadd et al[20] (protein description available in Table SII) were associated with CVD risk over 16 years of follow-up (n=12 657; n_{events}=1274).

First, we generated 109 Cox PH CVD risk models adjusted for ASSIGN. Each model was additionally adjusted for a different protein EpiScore. Two EpiScores failed to satisfy the PH assumption (Schoenfeld residual test P>0.05), and 6 EpiScores were not unique (proxied the concentration of the same protein). Of the remaining 101 protein EpiScores, 67 were significantly associated with CVD risk (P<0.05). After applying a conservative Bonferroni threshold for multiple testing (P<0.05/101=5.0×10^{-4}), 36 associations remained statistically significant.

Secondly, to understand whether protein EpiScores were associated with CVD risk beyond established biomarkers such as cardiac troponin, we included the concentration of cTnI as a covariate in the model along with ASSIGN, and we repeated the analysis. Of the 101

Table 1. Summary of Training, Test, and Full Data Sets. The Full Data Set Contains Related Individuals

	Training		Test		Full	
	Cases	Controls	Cases	Controls	Cases	Controls
n	658	6222	337	3322	1274	11 383
Time-to-event (years to onset or censoring)	7.0 (4.1–9.9)	11.8 (11.1–13.0)	4.8 (2.6–7.6)	11.8 (11.0–13.6)	6.8 (3.7–9.8)	11.8 (11.1–13.2)
Age, y	58.3 (50.8–62.6)	50.0 (40.8–58.8)	56.6 (51.6–60.0)	51.5 (43.9–57.6)	57.4 (51.0–62.2)	50.4 (41.5–58.2)
Sex, male	345 (52.4%)	2452 (39.4%)	165 (49.0%)	1219 (36.7%)	655 (51.4%)	4399 (38.6%)
SIMD, score/10	41.6 (22.2–53.3)	45.3 (26.3–55.1)	45.1 (20.2–54.9)	44.5 (22.6–54.8)	41.8 (21.6–54.0)	44.9 (25.5–54.9)
Family history of CHD/stroke, yes	443 (67.3%)	3171 (51.0%)	224 (66.5%)	1781 (53.6%)	862 (67.7%)	5881 (51.7%)
Diabetes, yes	19 (2.9%)	63 (1.0%)	16 (4.7%)	72 (2.2%)	43 (3.4%)	165 (1.4%)
Rheumatoid arthritis, yes	29 (4.4%)	140 (2.3%)	23 (6.8%)	110 (3.3%)	71 (5.6%)	281 (2.5%)
Nonsmoker, yes	534 (81.2%)	5306 (85.3%)	280 (83.1%)	2752 (82.8%)	1044 (81.9%)	9623 (84.5%)
Systolic blood pressure, mm Hg	142.1 (16.7)	130.8 (16.9)	140.3 (17.6)	130.3 (16.5)	141.5 (17.2)	130.6 (16.8)
Total cholesterol, mmol/L	5.4 (1.1)	5.2 (1.0)	5.3 (1.1)	5.3 (1.1)	5.4 (1.1)	5.2 (1.0)
HDL cholesterol, mmol/L	1.3 (1.1–1.6)	1.4 (1.2–1.7)	1.4 (1.1–1.6)	1.5 (1.2–1.8)	1.3 (1.1–1.6)	1.4 (1.2–1.7)
ASSIGN score	19 (12–29)	9 (4–18)	18 (11–28)	10 (5–17)	18 (11–28)	9 (4–17)

To make sure that members of the same family are not present across training and test data sets, any individuals in the training set who shared family ID with individuals from the test set were excluded from subsequent analyses (n=2118). For continuous variables with normal distributions, summary values are reported as mean (SD). Median (Q1-Q3) are given for continuous variables that do not follow a normal distribution. A number and a percentage of samples are reported for categorical variables.

ASSIGN indicates the cardiovascular risk score chosen for use by SIGN (Scottish Intercollegiate Guidelines Network) and Scottish Government Health Directorates; CHD, coronary heart disease; HDL, high-density lipoprotein; ID, identification number; and SIMD, Scottish Index of Multiple Deprivation.

aforementioned protein EpiScores, 65 were associated with CVD over and above the ASSIGN score and the concentration of cTnI ($P<0.05$; Figure 2). Thirty-three associations remained significant after correcting for multiple tests. Of the 65 protein EpiScores, higher levels of 41 were associated with an increased hazard of CVD (HR>1 and $P<0.05$). For example, elevated levels of CRP and MMP12 were associated with HR per SD of 1.23 (95% CI, 1.16–1.30; $P=9.2\times10^{-12}$) and 1.13 (95% CI, 1.06–1.22; $P=5.4\times10^{-4}$; Figure 3A), respectively. In contrast, higher levels of 24 protein EpiScores were associated with a decreased hazard of CVD (HR<1 and $P<0.05$). Examples of protein EpiScores belonging to this group include NOTCH1 (HR per SD, 0.84 [95% CI, 0.79–0.89]; $P=1.6\times10^{-9}$) and OMD (HR per SD, 0.87 [95% CI, 0.82–0.92]; $P=1.0\times10^{-6}$). The relationships between individual EpiScores and CVD risk have been visualized in the form of risk-over-time (Figure 3B), forest, and Kaplan Meier plots in an online R application (https://shiny.igc.ed.ac.uk/3d2c8245001b4e67875ddf2ee3fcbad2/).

As DNAm levels vary between different types of white blood cells, there is a concern that the associations that we observe may be influenced by cellular heterogeneity. To mitigate this potential effect, we incorporated estimated white blood cell proportions as covariates in the model adjusted for the concentration of cTnI and the ASSIGN score. In this model, 50 protein EpiScores were significantly associated with CVD risk ($P<0.05$). The comparison of HRs associated with protein EpiScores in each of the studied models can be found in Table SIII.

Finally, to learn whether individual protein EpiScore can augment CVD prediction beyond established biomarkers and clinical risk prediction tools, we calculated C-statistics for null and full models. While the null model was adjusted for ASSIGN and the concentration of cTnI (C-stat, 0.728), the full model also contained the studied protein EpiScore. Table 2 lists the top 10 associations that result in the greatest improvement in CVD risk prediction.

Composite Episcore for CVD Risk Prediction

To understand whether the abovementioned protein EpiScores can be used as biomarkers that add additional predictive value over and above typically used clinical risk scores (ASSIGN and SCORE2) and the concentration of cTnI, we generated a composite CVD EpiScore—a weighted linear combination of individual protein EpiScores. The score was trained using 2 modeling techniques: Cox PH Elastic Net and Random Survival Forest. There were 6880 records in the training set and 3659 records in the test set. The Elastic Net assigned nonzero coefficients to 45 of 109 protein EpiScores (Table SIV).

In a 10-year Elastic Net prediction analysis, the null model (containing age, sex, and ASSIGN) had an area under the receiver operating characteristic curve (AUC) of 0.719. The model with the CVD EpiScore increased the AUC to 0.723. The addition of cTnI to the null model resulted in an AUC of 0.721. The full model (null model+cTnI+CVD EpiScore) AUC was 0.724. Full output for the CVD models including C-statistics and a comparison with SCORE2 can be found in Tables V through VII. These analyses were a carbon copy of the aforementioned ASSIGN models—a null model (containing age, sex, and

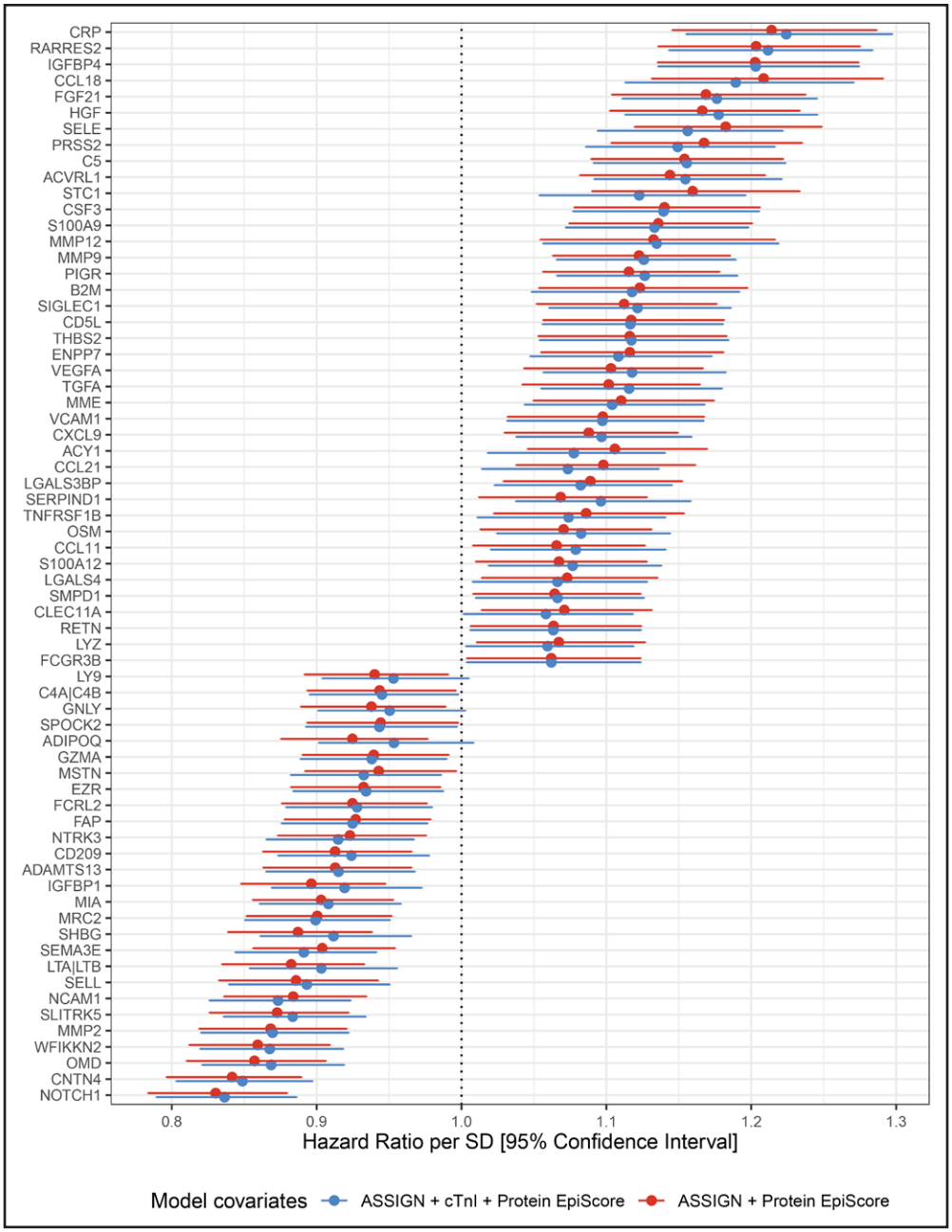

Figure 2. Associations between protein epigenetic scores (EpiScores) and incident cardiovascular disease.
Hazard ratios are plotted for the 67 significant associations (P<0.05) with 95% CI limits. Basic models were adjusted for ASSIGN (red), whereas full models included the ASSIGN score and concentration of cTnI (cardiac troponin I) as covariates (blue).

SCORE2) was compared with models with cTnI and the CVD EpiScore. The CVD EpiScore remained statistically significant after adjusting for the concentration of cTnI in models incorporating ASSIGN and SCORE2 (HR, 1.32; P=3.7×10^{-3} and HR, 1.36; P=1.4×10^{-3}, respectively).

Random Survival Forest–based analysis (see Methods) yielded similar results. The null model (as above) had an AUC of 0.719. Adding the CVD EpiScore to the null model increased the AUC to 0.721. The full model adjusted for CVD EpiScore and the concentration of cardiac troponin had an AUC of 0.723.

DISCUSSION

In this study, we describe 65 novel epigenetic biomarkers that are associated with long-term risk of CVD independently of a clinical risk prediction tool (ASSIGN) and the concentration of an established protein biomarker (cTnI). The most statistically significant EpiScores reflected concentrations of proteins involved in metabolic, immune, and developmental pathways. A weighted linear combination of protein EpiScores (the composite protein-CVD EpiScore) was significantly associated with CVD risk in

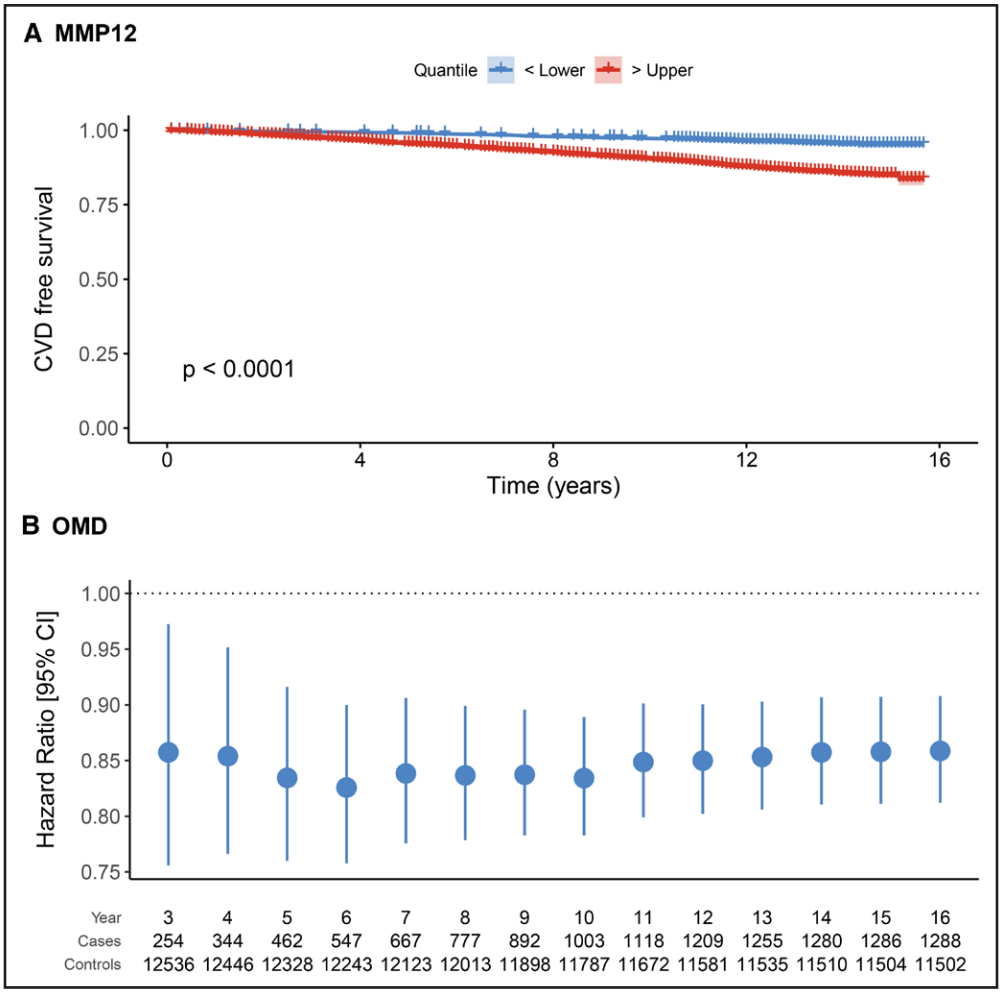

Figure 3. Changes in cardiovascular disease (CVD) free survival and CVD risk plotted for two selected protein EpiScores. **A**, Individuals with higher levels of MMP12 (>75th percentile) had shorter CVD-free survival when compared with those with lower levels of this EpiScore (<25th percentile). **B**, Hazard ratios (per SD of the EpiScore) and 95% CIs associated with the levels of OMD EpiScore plotted over time. At all examined time points, the association with CVD risk was significant (P<0.05).

models adjusted for ASSIGN. Although the score may be a useful addition to other omic features in future CVD risk prediction tools, at present, it is unlikely to be measured in a clinical setting.[23]

One previous study focused on how DNAm biomarkers improve CVD risk prediction.[24] Using time-to-event data and a panel of 60 blood DNAm biomarkers measured in an Italian cohort of 1803 individuals (295 cases), Cappozzo et al[24] trained a composite score for predicting short-term risk of CVD. In comparison, we focused on a more extensive panel of DNAm protein markers in addition to measured troponin levels. We also ran univariate analyses to identify individual proteins and protein classes that are associated with CVD. Furthermore, we developed 10-year prediction models (the prediction window for which both ASSIGN and SCORE2 are recommended) trained on more than double the number of cases.

Our findings suggest that individual protein EpiScores capture disease-specific biomarker signals relevant to CVD risk prediction. The relationships found between 65 protein EpiScores and incident CVD mirrored previously reported associations between CVD and measured protein concentrations. For example, elevated levels of CRP, a marker for systemic low-grade inflammation, have been associated with multiple age-related morbidities, including CVD.[25] MMP12 and OMD, in turn, are involved in maintaining the stability of atherosclerotic plaques. While MMP12 contributes to the growth and destabilization of plaques,[26] increased levels of OMD have been observed in macrocalcified plaques from asymptomatic patients.[27] Finally, multiple studies have demonstrated that NOTCH1 signaling protects the heart from CVD-induced myocardial damage. The Notch1 pathway is involved in neoangiogenesis and revascularization of a failing heart.[28] It limits the extent of ischemic injury,[28] reduces fibrosis,[29] and improves cardiac function.[30] Several protein EpiScores associated with CVD in our study, such as SELE and C5, have also been shown to be associated with stroke and ischemic heart disease in our previous work.[20]

Table 2. C-Statistics Calculated for Null and Full Protein EpiScore Models. Risk Was Ascertained Over 16 Years of Follow-Up. The Null Model Was Adjusted for ASSIGN and the Concentration of Cardiac Troponin I, While the Full Model Also Included a Studied EpiScore

EpiScore	C_{full}–C_{null}	Function
IGFBP4	0.0050	Metabolic/growth promoter
CRP	0.0050	Immune response
NTRK3	0.0046	Neural development/cell signaling
FGF21	0.0042	Metabolic
CSF3	0.0039	Immune response
HGF	0.0035	Growth factor/tissue regeneration
ACVRL1	0.0035	Vascular
CNTN4	0.0035	Cell adhesion/maintenance
PIGR	0.0034	Immune response
RARRES2	0.0032	Metabolic

EpiScore indicates epigenetic score.

Whereas some of the EpiScores reflect known protein-CVD associations, others reflect novel pathways. This includes, but is not limited to, PRSS2 and CNTN4. PRSS2, which encodes the digestive enzyme trypsin 2, has been mainly studied in the context of pancreatitis. However, recent studies provide evidence that trypsin can leak from the small intestine into the bloodstream and digest myocardial tissue during heart failure.[31] Trypsin-mediated degradation of heart tissue was also observed in cases of dilated cardiomyopathy following influenza A infection.[32] CNTN4, in turn, encodes a cell adhesion molecule implicated in the development of autism spectrum disorders.[33] Recent studies have shown that mutations in CNTN4 were associated with an elevated production of a prothrombotic agent called thromboxane A2 and an increased risk of cardiovascular events.[34]

The protein EpiScore that we trained for cTnI was not associated with the incidence of CVD. Therefore, we excluded it from composite CVD score generation. This highlights an important consideration in the development of multiomics biomarkers, as there are unlikely to be DNAm differences that associate with every blood protein. For example, the 109 protein EpiScores generated by Gadd et al[20] that we make use of in our study were extracted as the best-performing EpiScores from a total set of 953 proteins tested as potential outcomes. It is, therefore, not always possible to generate a meaningful protein EpiScore that reflects the protein biology. In the case of cardiac troponins, the elevations in circulating cTnI and cTnT are a result of a leakage of these proteins from the damaged heart muscle into the bloodstream.[35] As opposed to transcription, this process is not regulated by DNAm. Therefore, the methylation signal underlying an increased concentration of cardiac troponin in the bloodstream may be too weak to enable the generation of a meaningful EpiScore. This limitation may also extend to other proteins derived in the heart or other tissues involved in CVD onset. Nonetheless, the ability of a DNAm array to capture surrogate markers for hundreds of proteins—many of which are not routinely measured in the clinic—offers promise in the development of CVD biomarkers.

Strengths of this study include the precise timing of the CVD event through the electronic health records, the ability to generate a clinical risk predictor in a population cohort, and the large sample size for DNAm, which also permitted the splitting of the data into train/test sets to formally examine the improvement in risk prediction from our omics biomarkers.

Limitations to this work include the generalizability beyond a Scottish population. In this study, we trained and tested predictors in a Scottish cohort to augment the ASSIGN score. However, many of the protein EpiScores were trained in a German cohort (KORA [Cooperative Health Research in the Region Augsburg]) and projected to GS.[20] This suggests that the EpiScore biomarkers part-translate across European ancestry populations. Although the ASSIGN score is tailored to the Scottish population, we observed similar findings across all models when replacing it with SCORE2, which is widely used across Europe. To generalize the findings further, replication of the EpiScore associations with CVD (while adjusting for SCORE2) across other European ancestry populations is required.

CONCLUSIONS

In conclusion, we identified novel epigenetic signals that were associated with the incidence of CVD independently of ASSIGN and the concentration of cardiac troponin. The exploration of associations between protein EpiScores and CVD shed light on the etiology and molecular biology of the disease. As DNAm and proteins are assessed in increasingly large cohort samples, it will be possible to evaluate more precisely the potential gains in risk prediction, disease prevention, and any associated health economic benefits.

ARTICLE INFORMATION

Received April 22, 2023; accepted November 30, 2023.

Affiliations

Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer (A.D.C., D.A.G., Y.C., E.B., A.C., D.L.M., K.L.E., R.E.M.), Division of Psychiatry, Royal Edinburgh Hospital (A.M.M.), Edinburgh Clinical Research Facility, Western General Hospital (N.W., L.M.), and Usher Institute, Old Medical School (J.F.P.), The University of Edinburgh, United Kingdom. School of Psychology, University of Exeter, United Kingdom (R.M.W.). Institute of Cardiovascular and Medical Sciences, British Heart Foundation Glasgow Cardiovascular Research Centre, University of Glasgow, United Kingdom (P.W., N.S.).

Acknowledgments

The authors are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, health care assistants, and nurses. They would also like to thank Dr Robert Hillary for helpful feedback on the article and analyses.

Sources of Funding

This research was funded in whole, or in part, by the Wellcome Trust (104036/Z/14/Z, 108890/Z/15/Z, 220857/Z/20/Z, and 216767/Z/19/Z). For the purpose of open access, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission. A.D. Chybowska was supported by a Medical Research Council PhD Studentship in Precision Medicine with funding from the Medical Research Council Doctoral Training Program and the University of Edinburgh College of Medicine and Veterinary Medicine. D.A. Gadd was funded by the Wellcome Trust 4-Year PhD in Translational Neuroscience—Training the Next Generation of Basic Neuroscientists to Embrace Clinical Research (108890/Z/15/Z). Drs Bernabeu and Marioni were supported by the Alzheimer's Society major project grant AS-PG-19b-010. Generation Scotland (GS) received core support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006). Genotyping and DNA methylation profiling of the GS samples were performed by the Genetics Core Laboratory at the Clinical Research Facility, The University of Edinburgh, United Kingdom, and was funded by the Medical Research Council UK and the Wellcome Trust (references 104036/Z/14/Z, 220857/Z/20/Z, and 216767/Z/19/Z). The DNA methylation profiling and analysis were supported by Wellcome Investigator Award 220857/Z/20/Z and grant 104036/Z/14/Z (principal investigator: AM McIntosh) and through funding from National Alliance for Research on Schizophrenia & Depression (reference 27404; awardee: Dr D.M. Howard) and the Royal College of Physicians of Edinburgh (Sim Fellowship; awardee: Dr H.C. Whalley). All components of GS:SFHS (Generation Scotland: the Scottish Family Health Study), including the protocol and written study materials, have received formal and national ethical approval from the NHS Tayside Research Ethics Committee (Research Ethics Committee reference number 05/S1401/89). In addition, local approval has been obtained from the NHS Glasgow Research Ethics Committee and the NHS Glasgow and NHS Tayside Research and Development Offices, as required.

Disclosures

L. Murphy and Dr Marioni received a speaker fee from Illumina. Dr Marioni is an advisor to the Epigenetic Clock Development Foundation. D.A. Gadd and Dr Marioni received consultancy fees from Optima Partners. The other authors report no conflicts.

Supplemental Material

Supplemental Methods
Tables S1–S7
Figures S1–S3
References 36–46

REFERENCES

1. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ*. 2017;357:j2099. doi: 10.1136/bmj.j2099
2. van Staa TP, Gulliford M, Ng ESW, Goldacre B, Smeeth L. Prediction of cardiovascular risk using Framingham, ASSIGN and QRISK2: how well do they predict individual rather than population risk? *PLoS One*. 2014;9:e106455. doi: 10.1371/journal.pone.0106455
3. Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, Tzoulaki I, Lassale CM, Siontis GCM, Chiocchia V, Roberts C, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*. 2016;353:i2416. doi: 10.1136/bmj.i2416
4. Ho JE, Lyass A, Courchesne P, Chen G, Liu C, Yin X, Hwang S, Massaro JM, Larson MG, Levy D. Protein biomarkers of cardiovascular disease and mortality in the community. *J Am Heart Assoc*. 2018;7:e008108. doi: 10.1161/JAHA.117.008108
5. Hijazi Z, Lindbäck J, Alexander JH, Hanna M, Held C, Hylek EM, Lopes RD, Oldgren J, Siegbahn A, Stewart RAH, et al; ARISTOTLE and STABILITY Investigators. The ABC (age, biomarkers, clinical history) stroke risk score: a biomarker-based risk score for predicting stroke in atrial fibrillation. *Eur Heart J*. 2016;37:1582–1590. doi: 10.1093/eurheartj/ehw054
6. Wallentin L, Eriksson N, Olszowka M, Grammer TB, Hagström E, Held C, Kleber ME, Koenig W, März W, Stewart RAH, et al. Plasma proteins associated with cardiovascular death in patients with chronic coronary heart disease: a retrospective study. *PLoS Med*. 2021;18:e1003513. doi: 10.1371/journal.pmed.1003513
7. Gadd DA, Hillary RF, Kuncheva Z, Mangelis T, Admanit R, Gagnon J, Lin T, Ferber K, Runz H, Team BB, et al. Blood protein levels predict leading incident diseases and mortality in UK Biobank. 10.1101/2023.05.01.23288879.
8. Welsh P, Preiss D, Hayward C, Shah ASV, McAllister D, Briggs A, Boachie C, McConnachie A, Padmanabhan S, Welsh C, et al. Cardiac troponin T and troponin I in the general population. *Circulation*. 2019;139:2754–2764. doi: 10.1161/CIRCULATIONAHA.118.038529
9. Fridén V, Starnberg K, Muslimovic A, Ricksten S-E, Bjurman C, Forsgard N, Wickman A, Hammarsten O. Clearance of cardiac troponin T with and without kidney function. *Clin Biochem*. 2017;50:468–474. doi: 10.1016/j.clinbiochem.2017.02.007
10. Michielsen ECHJ, Wodzig WKWH, Van Dieijen-Visser MP. Cardiac troponin T release after prolonged strenuous exercise. *Sports Med*. 2008;38:425–435. doi: 10.2165/00007256-200838050-00005
11. Wu AHB. Release of cardiac troponin from healthy and damaged myocardium. *Front Lab Med*. 2017;1:144–150. doi: 10.1016/j.flm.2017.09.003
12. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev*. 2011;25:1010–1022. doi: 10.1101/gad.2037511
13. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*. 2008;9:465–476. doi: 10.1038/nrg2341
14. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet*. 2009;10:295–304. doi: 10.1038/nrg2540
15. Hop PJ, Luijk R, Daxinger L, van Iterson M, Dekkers KF, Jansen R, Heijmans BT, 't Hoen PAC, van Meurs J, Jansen R, et al. Genome-wide identification of genes regulating DNA methylation using genetic anchors for causal inference. *Genome Biol*. 2020;21:220.
16. Moldoveanu AI, Shephard RJ, Shek PN. Exercise elevates plasma levels but not gene expression of IL-1β, IL-6, and TNF-α in blood mononuclear cells. *J Appl Physiol*. 2000;89:1499–1504. doi: 10.1152/jappl.2000.89.4.1499
17. Koenig W, Sund M, Fröhlich M, Löwel H, Hutchinson WL, Pepys MB. Refinement of the association of serum C-reactive protein concentration and coronary heart disease risk by correction for within-subject variation over time: the MONICA Augsburg studies, 1984 and 1987. *Am J Epidemiol*. 2003;158:357–364. doi: 10.1093/aje/kwg135
18. Liu Y, Buil A, Collins BC, Gillet LCJ, Blum LC, Cheng LY, Vitek O, Mouritsen J, Lachance G, Spector TD, et al. Quantitative variability of 342 plasma proteins in a human twin population. *Mol Syst Biol*. 2015;11:786. doi: 10.15252/msb.20145728
19. Stevenson AJ, McCartney DL, Hillary RF, Campbell A, Morris SW, Bermingham ML, Walker RM, Evans KL, Boutin TS, Hayward C, et al. Characterisation of an inflammation-related epigenetic score and its association with cognitive ability. *Clin Epigenetics*. 2020;12:113. doi: 10.1186/s13148-020-00903-8
20. Gadd DA, Hillary RF, McCartney DL, Zaghlool SB, Stevenson AJ, Cheng Y, Fawns-Ritchie C, Nangle C, Campbell A, Flaig R, et al. Epigenetic scores for the circulating proteome as tools for disease prediction. Lo YD, Ferrucci L, eds. *eLife*. 2022;11:e71802. doi: 10.7554/eLife.71802
21. Estimate the risk-ASSIGN Score. Accessed April 14, 2022. https://www.assign-score.com/estimate-the-risk/visitors/
22. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, Dominiczak AF, Fitzpatrick B, Ford I, Jackson C, et al. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med Genet*. 2006;7:74. doi: 10.1186/1471-2350-7-74
23. Wagner W. How to translate DNA methylation biomarkers into clinical practice. *Front Cell Dev Biol*. 2022;10:854797. doi: 10.3389/fcell.2022.854797
24. Cappozzo A, McCrory C, Robinson O, Freni Sterrantino A, Sacerdote C, Krogh V, Panico S, Tumino R, Iacoviello L, Ricceri F, et al. A blood DNA methylation biomarker for predicting short-term risk of cardiovascular events. *Clin Epigenetics*. 2022;14:121. doi: 10.1186/s13148-022-01341-4
25. Sharif S, Van der Graaf Y, Cramer MJ, Kapelle LJ, de Borst GJ, Visseren FLJ, Westerink J, van Petersen R, Dinther BGF, Algra A, et al. Low-grade inflammation as a risk factor for cardiovascular events and all-cause mortality in patients with type 2 diabetes. *Cardiovasc Diabetol*. 2021;20:220.
26. Goncalves I, Bengtsson E, Colhoun HM, Shore AC, Palombo C, Natali A, Edsfeldt A, Dunér P, Fredrikson GN, Björkbacka H, et al; SUMMIT Consortium. Elevated plasma levels of MMP-12 are associated with atherosclerotic burden and symptomatic cardiovascular disease in subjects with type 2 diabetes. *Arterioscler Thromb Vasc Biol*. 2015;35:1723–1731. doi: 10.1161/ATVBAHA.115.305631
27. Gonçalves I, Oduor L, Matthes F, Rakem N, Meryn J, Skenteris NT, Aspberg A, Orho-Melander M, Nilsson J, Matic L, et al. Osteomodulin gene expression is associated with plaque calcification, stability, and fewer cardiovascular events in the CPIP cohort. *Stroke*. 2022;53:e79–e84. doi: 10.1161/STROKEAHA.121.037223
28. Li Y, Hiroi Y, Ngoy S, Okamoto R, Noma K, Wang CY, Wang H-W, Zhou Q, Radtke F, Liao R, et al. Notch1 in bone marrow–derived cells mediates

cardiac repair after myocardial infarction. *Circulation*. 2011;123:866–876. doi: 10.1161/CIRCULATIONAHA.110.947531
29. Ferrari R, Rizzo P. The Notch pathway: a novel target for myocardial remodelling therapy? *Eur Heart J*. 2014;35:2140–2145. doi: 10.1093/eurheartj/ehu244
30. Gude NA, Emmanuel G, Wu W, Cottage CT, Fischer K, Quijada P, Muraski JA, Alvarez R, Rubio M, Schaefer E, et al. Activation of notch-mediated protective signaling in the myocardium. *Circ Res*. 2008;102:1025–1035. doi: 10.1161/CIRCRESAHA.107.164749
31. Courelli V, Ahmad A, Ghassemian M, Pruitt C, Mills PJ, Schmid-Schönbein GW. Digestive enzyme activity and protein degradation in plasma of heart failure patients. *Cell Mol Bioeng*. 2021;14:583–596. doi: 10.1007/s12195-021-00693-w
32. Pan HY, Sun HM, Xue LJ, Pan M, Wang YP, Kido H, Zhu JH. Ectopic trypsin in the myocardium promotes dilated cardiomyopathy after influenza a virus infection. *Am J Physiol Heart Circ Physiol*. 2014;307:H922–H932. doi: 10.1152/ajpheart.00076.2014
33. Zhang SQ, Fleischer J, Al-Kateb H, Mito Y, Amarillo I, Shinawi M. Intragenic CNTN4 copy number variants associated with a spectrum of neurobehavioral phenotypes. *Eur J Med Genet*. 2020;63:103736. doi: 10.1016/j.ejmg.2019.103736
34. McCarthy N, Vangjeli C, Surendran P, Treumann A, Rooney C, Ho E, Sever P, Thom S, Hughes A, Munroe P, et al. Genetic variants in PPARGC1B and CNTN4 are associated with thromboxane A2 formation and with cardiovascular event free survival in the Anglo-Scandinavian Cardiac Outcomes Trial (ASCOT). *Atherosclerosis*. 2018;269:42–49. doi: 10.1016/j.atherosclerosis.2017.12.013
35. Daubert MA, Jeremias A. The utility of troponin measurement to detect myocardial infarction: review of the current findings. *Vasc Health Risk Manag*. 2010;6:691–699. doi: 10.2147/vhrm.s5306
36. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, Deary IJ, MacIntyre DJ, Campbell H, McGilchrist M, et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol*. 2013;42:689–700. doi: 10.1093/ije/dys084
37. Welsh P, Preiss D, Shah ASV, McAllister D, Briggs A, Boachie C, McConnachie A, Hayward C, Padmanabhan S, Welsh C, et al. Comparison between high-sensitivity cardiac troponin T and cardiac troponin I in a large general population cohort. *Clin Chem*. 2018;64:1607–1616. doi: 10.1373/clinchem.2018.292086
38. Cheng Y, Gadd DA, Gieger C, Monterrubio-Gómez K, Zhang Y, Berta I, Stam MJ, Szlachetka N, Lobzaev E, Wrobel N, et al. Development and validation of DNA methylation scores in two European cohorts augment 10-year risk prediction of type 2 diabetes. *Nat Aging* 2023;3:450–458. doi: 10.1038/s43587-023-00391-4
39. Faqs-ASSIGN Score. Accessed April 14, 2022. https://www.assign-score.com/faqs/
40. Woodward M, Brindle P, Tunstall-Pedoe H., for the SIGN group on risk estimation*. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart*. 2005;93:172–176. doi: 10.1136/hrt.2006.108167
41. SCORE2 Working Group and ESC Cardiovascular Risk Collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J*. 2021;42:2439–2454. doi: 10.1093/eurheartj/ehab309
42. SCORE2 risk calculator for low-risk regions. Accessed October 6, 2022. https://heartscore.escardio.org/Calculate/quickcalculator.aspx?model=low
43. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. 1st ed. Springer New York; 2013. doi: 10.1007/978-1-4757-3294-8
44. Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Statist Soft*. 2010;33:1–22.
45. Lin DY. On the Breslow estimator. *Lifetime Data Anal*. 2007;13:471–480. doi: 10.1007/s10985-007-9048-y
46. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf*. 2011;12:77. doi: 10.1186/1471-2105-12-77