ESC European Society of Cardiology

Europace (2023) **25**, 1–14
https://doi.org/10.1093/europace/euad320

TRANSLATIONAL RESEARCH

A proteomic analysis of atrial fibrillation in a prospective longitudinal cohort (AGES-Reykjavik study)

Thorarinn Jonmundsson [1,2], Anna E. Steindorsdottir [2], Thomas R. Austin [3,4], Elisabet A. Frick [1,2], Gisli T. Axelsson [1,2], Lenore Launer [5], Bruce M. Psaty[3,4,6], Joseph Loureiro [7], Anthony P. Orth [8], Thor Aspelund [1,2], Valur Emilsson [1,2], James S. Floyd [3,4†], Lori Jennings [7†], Vilmundur Gudnason[1,2*†], and Valborg Gudmundsdottir [1,2*†]

[1]Icelandic Heart Association, Holtasmari 1, Kopavogur 201, Iceland; [2]Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland; [3]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA; [4]Department of Epidemiology, University of Washington, Seattle, WA, USA; [5]Laboratory of Epidemiology and Population Sciences, Intramural Research Program, National Institute on Aging, Bethesda, MD, USA; [6]Department of Health Systems and Population Health, University of Washington, Seattle, WA, USA; [7]Novartis Biomedical Research, Cambridge, MA, USA; and [8]Novartis Biomedical Research, San Diego, CA, USA

Received 9 August 2023; accepted after revision 6 October 2023; online publish-ahead-of-print 15 November 2023

Aims Atrial fibrillation (AF) is associated with high risk of comorbidities and mortality. Our aim was to examine causal and predictive relationships between 4137 serum proteins and incident AF in the prospective population-based Age, Gene/Environment Susceptibility-Reykjavik (AGES-Reykjavik) study.

Methods and results The study included 4765 participants, of whom 1172 developed AF. Cox proportional hazards regression models were fitted for 4137 baseline protein measurements adjusting for known risk factors. Protein associations were tested for replication in the Cardiovascular Health Study (CHS). Causal relationships were examined in a bidirectional, two-sample Mendelian randomization analysis. The time-dependent area under the receiver operating characteristic curve (AUC)-statistic was examined as protein levels and an AF-polygenic risk score (PRS) were added to clinical risk models. The proteomic signature of incident AF consisted of 76 proteins, of which 63 (83%) were novel and 29 (38%) were replicated in CHS. The signature included both N-terminal prohormone of brain natriuretic peptide (NT-proBNP)-dependent (e.g. CHST15, ATP1B1, and SVEP1) and independent components (e.g. ASPN, AKR1B, and LAMA1/LAMB1/LAMC1). Nine causal candidates were identified (TAGLN, WARS, CHST15, CHMP3, COL15A1, DUSP13, MANBA, QSOX2, and SRL). The reverse causal analysis suggested that most AF-associated proteins were affected by the genetic liability to AF. N-terminal prohormone of brain natriuretic peptide improved the prediction of incident AF events close to baseline with further improvements gained by the AF-PRS at all time points.

Conclusion The AF proteomic signature includes biologically relevant proteins, some of which may be causal. It mainly reflects an NT-proBNP-dependent consequence of the genetic liability to AF. N-terminal prohormone of brain natriuretic peptide is a promising marker for incident AF in the short term, but risk assessment incorporating a PRS may improve long-term risk assessment.

* Corresponding authors. Tel: +354 535 1800, *E-mail address*: valborg@hjarta.is (Va.G.); *E-mail address*: v.gudnason@hjarta.is (Vi.G.)

† The last four authors contributed equally to the study.

Structured Graphical Abstract

Key question

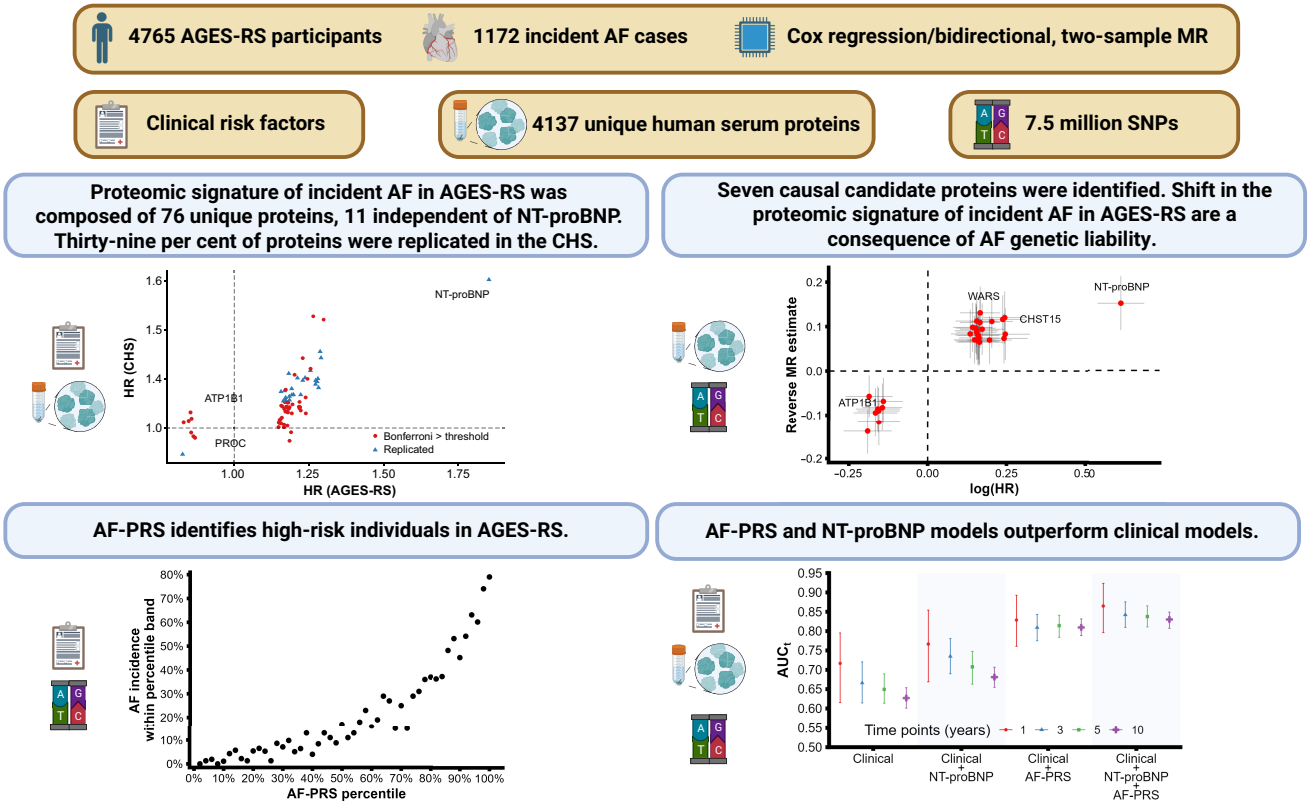
What is the composition of the proteomic signature for incident atrial fibrillation (AF) in the Age, Gene/Environment Susceptibility-Reykjavik study (AGES-RS), and which processes does it reflect? Is the change in protein serum levels a cause or consequence of the disease onset? What does this imply for prediction of incident AF?

Key findings

Seventy-six unique proteins were associated with incident AF, 11 of which are independent of N-terminal prohormone of brain natriuretic peptide (NT-proBNP). Seven putative causal candidate proteins were identified. The atrial fibrillation-polygenic risk score (PRS) achieves good separation between high- and low-risk groups for incident AF. The AF-PRS and NT-proBNP improve prediction over risk factors.

Take-home message

The proteomic signature in AGES-RS largely reflects the effects of NT-proBNP and a response to the genetic liability of AF. When NT-proBNP is accounted for, novel protein associations are revealed, and prediction of events close to baseline is improved. A risk model including AF-PRS and NT-proBNP provides the most promising and stable classification across multiple time points.



AGES-RS, Age, Gene/Environment Susceptibility-Reykjavik study; MR, Mendelian randomization; AF, atrial fibrillation; CHS, Cardiovascular Health Study; PRS, polygenic risk score.

Keywords Proteomics • Atrial fibrillation • Mendelian randomization • Polygenic risk score • Prediction • NT-proBNP

What's new?

- The proteomic signature of incident atrial fibrillation (AF) in the Age, Gene/Environment Susceptibility-Reykjavik study (AGES-RS) consisted of 76 unique proteins. Of those, 63 (83%) were novel and 29 (38%) were replicated successfully in the Cardiovascular Health Study (CHS) cohort. Furthermore, 11 protein associations with AF were independent of N-terminal prohormone of brain natriuretic peptide (NT-proBNP).
- Seven putative causal candidate proteins for AF were identified.
- Comparison of proteins and polygenic risk scores in AGES-RS showed that the AF-polygenic risk score (PRS) and NT-proBNP improved prediction over clinical risk factors.

Introduction

Atrial fibrillation (AF) is the most prevalent arrhythmia and is associated with a high risk of comorbidities, such as stroke or heart failure, and mortality.[1] The economic burden of AF has been estimated to be considerable[2,3] and will likely worsen as the incidence of AF is projected to increase significantly in the coming decades.[4] Consequently, AF is an important and active area of study and, despite its complexities, has seen advances in fields such as molecular biology[5] and genetics,[6] as well as emerging novel principles such as atrial inflammatory signalling.[7]

Recent advances in high-throughput protein measurements and large-scale genome-wide association studies (GWAS) have sparked interest in identifying novel biomarkers for AF and causal proteins that may provide novel therapeutic targets. A clinical risk model for incident AF developed by the CHARGE consortium,[8] which included established risk factors associated with development of AF,[1,9] established a benchmark against which subsequent studies for new risk factors have compared. Examples include proteomic studies conducted in the ARIC[10] and Framingham[11] cohorts which revealed several proteins associated with incident AF, and a recent study which highlighted the potential of polygenic risk scores (PRSs) and N-terminal prohormone of brain natriuretic peptide (NT-proBNP/NBBP) as predictors for incident AF in patients with cardiovascular conditions.[12] Additionally, recent Mendelian randomization (MR) studies have identified several proteins with a potential causal role in AF.[13,14] These studies were conducted independently of any observational data, while no causal estimation was reported in the observational studies.[10,11]

To unify and expand upon the previously established literature, we examined if serum levels of 4137 proteins measured by the SOMAscan platform consisting of 4782 slow off-rate modified aptamers (SOMAmers) were related to incident AF in the Age, Gene/Environment Susceptibility-Reykjavik Study (AGES-RS),[15] a prospective and population-based cohort. All proteins found to be significantly associated with incident AF were tested for replication in the Cardiovascular Health Study[16] (CHS) and examined for biological relevance with respect to cardiac function and structure. We extended these results further by performing a bidirectional, two-sample MR analysis[17] to assess both potential causation and reverse causation of the serum proteins associated with incident AF and evaluated the relationship between an AF-PRS and the serum proteome. Finally, we examined the predictive performance of the top-associated protein [NT-proBNP/natriuretic peptide B (NPPB)] together with an AF-PRS and clinical risk factors for incident AF at different time points in our population-based cohort.

Methods

Study population

The AGES-RS is a population-based and prospective study comprised of men and women born between 1907 and 1935 ($n = 5764$) and recruited from surviving participants ($n = 11\,549$) of the Reykjavik study (established 1967).[15] The AGES-RS participants underwent a comprehensive assessment for age-related diseases at baseline (2002–2006) and a second follow-up visit ($n = 3411$, 2007–2011). The follow-up period of this study was defined as the AGES-RS baseline date of entry until 19 March 2019. Atrial fibrillation cases were identified with electrocardiograms (ECGs) performed at the AGES-RS visits and International Classification of Diseases (ICD) codes from the hospitalization records (ICD9CM 427.3, 427.31, 427.32, or ICD10 I48 in any position) of the National University Hospital of Iceland. Participants with prevalent AF at baseline, incomplete clinical data, or unmeasured serum proteins were excluded from the study. Remaining participants were followed until the first occurrence of AF, death, or the end of the study. Death was treated as a censoring event unless cause of death was AF.

Data collected at baseline on anthropometry [height, weight, and body mass index (BMI)], lifestyle (smoking), physiology [blood pressure, heart rate, lipids, fasting glucose, insulin, HbA1c, C-reactive protein (CRP), and estimated glomerular filtration rate (eGFR)], ECG (PR interval, QRS interval, and QT interval), prevalent comorbidities (diabetes, myocardial infarction, and heart failure), and medication use (antihypertensive, antiarrhythmic, anticoagulation, diabetes, and statins) were summarized for incident AF and non-cases, and differences were tested with the Wilcoxon signed-rank test or *F*-test. The AGES-RS study was approved by the NBC in Iceland (approval number VSN-00-063), the National Institute on Aging (NIA) Intramural Institutional Review Board, and the Data Protection Authority in Iceland.

Serum protein measurements and genetic data

The serum protein levels of 4137 human proteins in the AGES-RS cohort were quantified with 4782 SOMAmer using the SomaScan proteomic profiling platform (Novartis V3-5K). The quantification process and its subsequent Box–Cox transformation have been described.[18,19] Single-nucleotide polymorphisms (SNPs) were genotyped with the Illumina hu370CNV array and the Illumina Infinium Global Screening Array. Both arrays were imputed against the Haplotype Reference Consortium imputation panel r1.1. After quality control, 7 506 463 variants for 5368 individuals were available for analysis, as previously described.[20]

Identification of a proteomic signature of incident atrial fibrillation in AGES-RS

Protein measurements were added, one at a time, to two Cox proportional hazards (PH) regression models (Models 1 and 2), utilizing the full follow-up time. Model 1 included CHARGE-AF risk factors[8] (age, height, weight, systolic and diastolic blood pressure, smoking status, antihypertensive medication use, diabetes status, and previous history of heart failure and myocardial infarction), sex, and eGFR as covariates. Model 2 additionally included the aptamer serum levels for NT-proBNP (gene symbol NPPB), in order to identify NPPB-independent protein signals (*Figure 1*). Protein HRs should be interpreted as the change in expected risk per unit increase of standard deviation (SD). The PH assumption (i.e. whether the HRs were constant over time) of the models was assessed with the cox.zph function from the survival R package.[21]

Incident AF protein associations reaching study-wide significance (Model 1, $P < 0.05/4782$; Model 2, $P < 0.05/4781$) were (i) compared with previously reported associations from two recently published studies[10,11] and (ii) tested for replication in the CHS.[16] All four cohorts resemble each other with respect to recruitment strategy and baseline characteristics (see Supplementary material online, *Tables S1* and *S2*) and use the same aptamer-based affinity proteomics technology to measure circulating protein levels. Proteins were measured in CHS with the 5 K ($n = 2944$) and 7 K ($n = 361$) SomaScan assays, where scaling factors provided by SomaLogic were applied to analyse the data jointly.

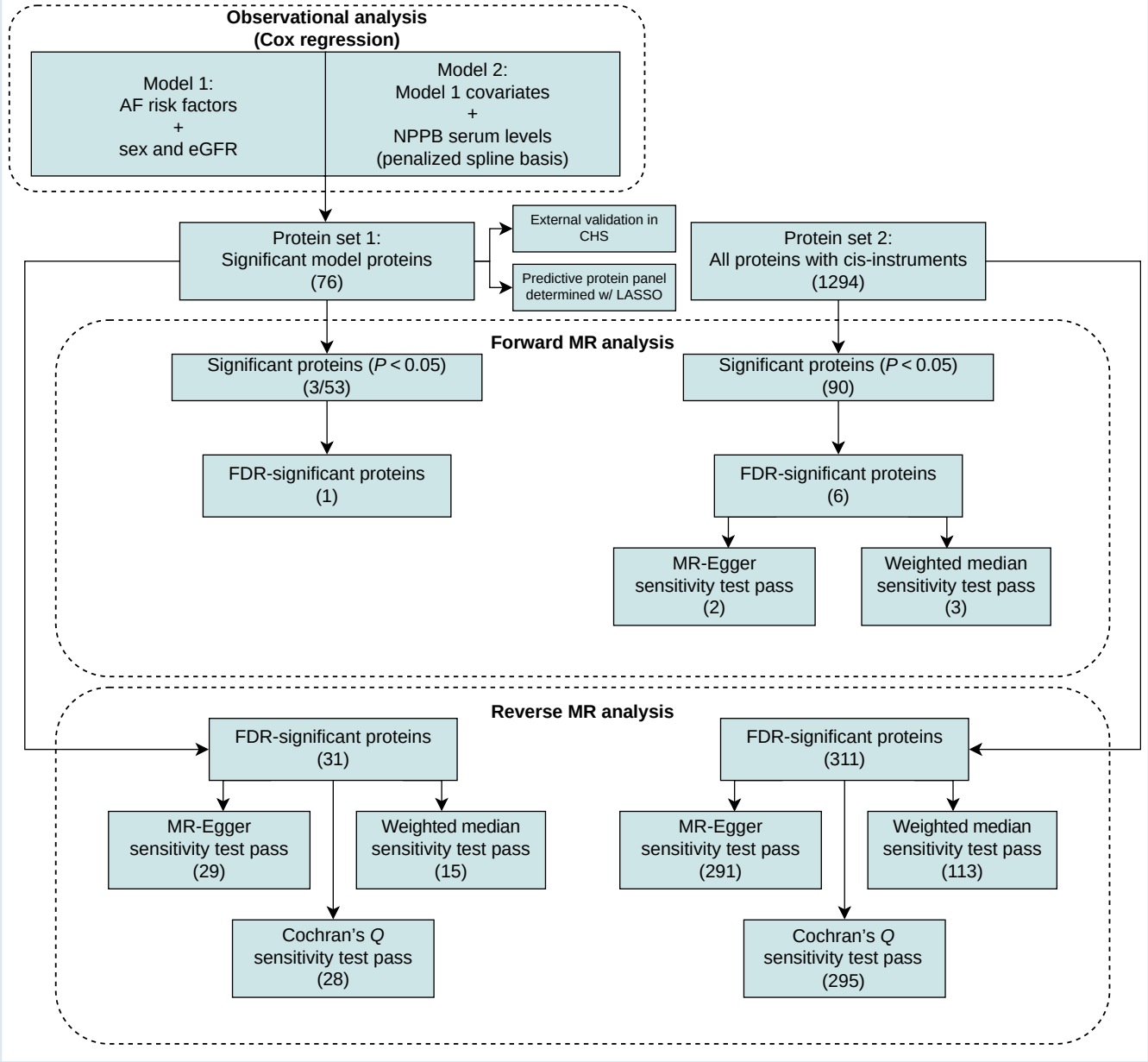


Figure 1 Flowchart of analyses performed to identify and validate the protein signature of incident AF in AGES and the subsequent bidirectional MR analysis.

Causal inference

A bidirectional, two-sample MR analysis[17] was performed to assess the putative causal relationship between proteins and AF (*Figure 1*). Summary statistics on the genetic risk of AF were obtained from an external GWAS composed of 60 620 AF cases and 970 216 controls.[22]

In the forward MR analysis (exposure, serum protein levels; outcome, AF), genetic instruments (SNPs which fulfil the assumptions of MR[23]; Supplementary Methods) for proteins were defined as cis-SNPs within 500 kb up- and downstream of the protein-encoding gene targeted by its respective aptamer, which were clumped based on linkage disequilibrium (LD) ($r^2 \geq 0.2$; kb window ± 500 kb) with PLINK v1.9[24] and filtered for cis-window significance ($P < 0.05$/#SNPs within cis-region prior to clumping) and strength (F-statistic > 10). Protein-associated SNPs missing in the AF GWAS were replaced by proxy variants ($r^2 > 0.8$) when possible. In the reverse MR analysis (exposure, AF; outcome, serum protein levels), genetic instruments for AF were defined as genome-wide significant SNPs ($P < 5 \times 10^{-8}$) after clumping (LD: $r^2 > 0.01$, kb window: ± 10 000 kb up- and downstream of variant position). Furthermore, SNPs within the cis-region for a given aptamer were removed in the reverse analysis to prevent the genetic liability of AF being expressed through processes directly dependent on the cis-SNPs. As the forward MR analysis was restricted to a neighbourhood of the protein-encoding gene, a less stringent LD threshold was chosen for a more inclusive SNP selection[25] (see Supplementary Methods). The resulting sets of SNPs for each MR analysis were harmonized with the TwoSampleMR R package.[26]

Single- and multi-SNP causal estimates were computed with Wald's ratio estimator and the generalized weighted least squares (GWLS) estimator, respectively. The GWLS estimator was appropriate for both MR analyses as it accounts for the correlation structure of the cis-region (forward MR) and reduces to the standard inverse variance weighted (IVW) estimator when SNPs are independent (reverse MR).[25,27] Forward MR causal estimates can be interpreted as the effect on the

probability of AF per unit increase in genetically predicted serum protein levels. Similarly, reverse MR causal estimates represent the effect on protein serum levels per 1% increase in the genetically predicted probability of AF.[28]

Significant causal estimates [Benjamini–Hochberg false discovery rate (FDR) < 0.05] based on at least three SNPs were subjected to a sensitivity analysis comprised of MR-Egger[23] and weighted median[29] (WM) estimation, modified to account for dependence when necessary (see Supplementary Methods). Reverse MR causal estimates were tested for heterogeneity with Cochran's Q[29] and locus-dependent significance in a leave-one-out locus sensitivity analysis (see Supplementary Methods).

Atrial fibrillation-polygenic risk score

The AF-PRS was computed for AGES-RS participants in a similar manner as described by Khera *et al.*[30] with PLINK v1.9[24] using summary statistics from the AF GWAS.[22] Sets of SNPs generated from all pairwise combinations of LD thresholds $r^2 = (0.2, 0.4, 0.6, 0.8)$ and P values (5×10^{-8}, 5×10^{-6}, 5×10^{-4}, 5×10^{-2}, 5×10^{-1}) within a ±250 kB window were considered and narrowed down to a single candidate with bootstrap resampling (see Supplementary Methods). Protein measurements were regressed on the candidate AF-PRS (see Supplementary Methods) and compared with the observational and causal protein analyses. Gene Set Enrichment Analysis (GSEA) of Gene Ontology terms was performed on the proteins associated with the AF-PRS using the R package fgsea.[31]

Prediction

Cumulative/dynamic receiver operating characteristic (ROC) curve estimation[32,33] was performed with the survivalROC R package[34] to calculate the time-dependent AUC[35] statistic (AUC_t). Full follow-up time was used to assess prediction of incident AF at different time points. Model 1 (CHARGE-AF risk factors, sex, and eGFR) was refitted to include all significant incident AF proteins identified in the observational analysis, in addition to a LASSO penalty under a bootstrap resampling schema (see Supplementary Methods). Proteins were retained for further analysis if they increased the *c*-statistic at full follow-up time and were selected in >70% of bootstrap iterations. For the selected proteins, their AUC_t statistic for events occurring within 10 years was compared with that of the CHARGE-AF risk factors (Model 1).

Finally, the AUC_t statistic was calculated for events occurring within 1, 3, 5, or 10 years for multiple models of interest. All models included age and sex as covariates, and the gain in predictive performance (as measured by the AUC_t statistic) was evaluated iteratively as combinations of CHARGE-AF risk factors, NPPB serum levels, and AF-PRS were added as predictors. Percentile intervals (2.5–97.5%) for each AUC_t statistic were constructed with 500 iterations of a bootstrap resampling schema (see Supplementary Methods).

Results

Cohort characteristics

Of 5457 AGES-RS participants with available protein and genetic data, 4765 remained after participants with prevalent AF ($n = 507$) or incomplete clinical data ($n = 185$) had been excluded from the study (see Supplementary material online, *Figure S1A*). In total, 1172 of 4765 study participants experienced an AF event (24.64 cases per 1000 patient-years), 873 of which did so within 10 years. The median follow-up for the non-case and incident group was 7.2 years and 12.4 years, respectively (see Supplementary material online, *Figure S1B*). The complete baseline characteristics of the AGES-RS cohort are shown in *Table 1*.

Serum proteins associated with incident atrial fibrillation

When Model 1 was fitted to the data, 65 unique proteins (corresponding to 76 aptamers) were significantly ($P < 0.05/4782$) associated with incident AF (*Figure 2A and B*, Supplementary material online, *Table S3*). The strongest association was observed for NPPB (NT-proBNP, $P = 2.22 \times 10^{-57}$), for which an increase of one-unit SD corresponded to an 86% increase in the expected hazard [hazard ratio (HR) = 1.86]. Among the top AF-associated proteins were some involved in cardiac conductivity (ATP1B1), vasculature development (NPPB, ANGPT2, SVEP1, TNNI3, VEGFD, ESM1, and WARS) and extracellular matrix (SPON1, POSTN, THBS2, and EPYC). Of the 65 proteins, 52 were novel findings (see Supplementary material online, *Table S4*) as 13 had been previously associated with AF in the ARIC or Framingham cohorts,[10,11] with directionally consistent HRs (see Supplementary material online, *Table S5*). All 65 proteins (74/76 aptamers) were assayed in CHS and could be included in a replication analysis (see Supplementary material online, *Table S4*). Of those 65, 24 (37%) proteins passed the replication threshold ($P < 0.05/74$) (see Supplementary material online, *Table S4*; *Figure 2C*). Furthermore, 15/52 (29%) of the novel protein associations were replicated successfully in CHS. All proteins replicated in CHS had a direction of effect consistent with their AGES-RS counterparts. Lastly, 26 of 60 proteins previously associated with incident AF in the ARIC and Framingham cohorts were replicated in AGES-RS (ARIC, $P < 0.05/60$; Framingham, $P < 0.05/3$; Supplementary material online, *Table S6*).

Natriuretic peptide B–independent serum proteins associated with incident atrial fibrillation

As NPPB had the strongest association with incident AF and was significant in all studies (see Supplementary material online, *Tables S3–S5*), we investigated how a model including NPPB as a covariate (Model 2) affected other protein associations. A penalized spline basis was specified for NPPB to address its PH violation (see Supplementary material online, *Tables S7 and S8*). Of the 65 unique proteins identified in Model 1, only CRP and TNNI3 remained significant ($P < 0.05/75$) in Model 2 (*Figure 3*, Supplementary material online, *Table S3*). As in Model 1, elevated serum levels of these proteins were associated with a higher risk of developing AF (TNNI3: HR = 1.12, $P = 9.29 \times 10^{-5}$; CRP: HR = 1.12, $P = 3.92 \times 10^{-4}$). Thus, for most of the serum proteins identified by Model 1, their association to incident AF is NPPB dependent.

When Model 2 was fit to the entire protein panel, 11 unique proteins (corresponding to 12 aptamers) were significantly ($P < 0.05/4781$) associated with incident AF (AKR1B1, ART3, ASPN, IGDCC4, LAMA1/LAMB1/LAMC1, LRRC24, LTB4R, PIANP, RASA1, UST, and WFIKKN2; *Figure 3*, Supplementary material online, *Table S3*). All proteins were protective for incident AF (see Supplementary material online, *Table S3*) and novel to the analysis as none reached study-wide significance in Model 1. Hence, these 11 proteins reflect NPPB-independent factors which affect incident AF but were masked by NPPB. Five of these associations (45%) for four unique proteins (ART3, IGDCC4, PIANP, WFIKKN2) were successfully replicated ($P < 0.05/12$) in CHS. All protein associations were directionally consistent with their AGES-RS counterparts (see Supplementary material online, *Table S4*).

Putative causal proteins for atrial fibrillation

In total, 76 unique proteins (corresponding to 88 aptamers) were significantly associated with incident AF in AGES-RS in Models 1 and 2. Of those, 53 had cis-acting genetic variants associated with their serum levels (see Supplementary material online, *Table S9*) and could be tested for causality. The forward MR analysis yielded three proteins with support ($P < 0.05$) for causality: CHST15, TAGLN, and WARS (*Figure 4A*, Supplementary material online, *Figures S2–S4* and *Table S3*), of which TAGLN remained significant after adjusting for multiple comparisons (FDR < 0.05). As TAGLN had a single cis-acting variant, it could not

Table 1 Baseline characteristics of AGES-RS cohort

Variable	Non-case	Incident AF	*P*	Total
Demographics				
N	3593 (75.4%)	1172 (24.6%)	NA	4765
Age (years)	75 (72–80)	76 (73–81)	4.70E−06	76 (72–80)
Women	2163 (60.2%)	662 (56.5%)	2.70E−02	2825 (59.3%)
Follow-up (years)	12.4 (7.4–14.5)	7.2 (4.4–10.2)	1.70E−124	10.5 (6.2–14.1)
Anthropometry				
Height (cm)	165.6 (159.7–173.3)	166.9 (160–175.1)	8.30E−04	165.8 (159.8–173.7)
Weight (kg)	73.4 (65–84)	75.8 (66.9–85.8)	4.20E−07	74.1 (65.4–84.5)
BMI (kg/m^2)	26.5 (23.9–29.4)	27.1 (24.5–29.9)	1.90E−04	26.7 (24.1–29.5)
Lifestyle				
Smoker (current)	466 (13%)	123 (10.5%)	2.90E−02	589 (12.4%)
Physiological				
Systolic blood pressure (mmHG)	140 (128–154)	142 (130–158)	3.40E−05	141 (128–155)
Diastolic blood pressure (mmHG)	74 (68–80)	73 (68–80)	>0.05	74 (68–80)
Heart rate (bpm)	66 (59–73)	64 (57–72)	1.20E−05	65 (58–73)
eGFR (mL/min/1.73 m^2)	64.5 (54.8–75.3)	64.3 (52.5–74.8)	>0.05	64.4 (54.4–75.3)
Total cholesterol (mmol/L)	5.7 (4.9–6.4)	5.6 (4.9–6.3)	3.10E−03	5.6 (4.9–6.4)
HDL cholesterol (mmol/L)	1.5 (1.3–1.9)	1.5 (1.2–1.9)	>0.05	1.5 (1.3–1.9)
LDL cholesterol (mmol/L)	3.5 (2.8–4.2)	3.5 (2.8–4.1)	1.70E−02	3.5 (2.8–4.2)
Triglycerides (mmol/L)	1 (0.8–1.4)	1 (0.8–1.4)	>0.05	1 (0.8–1.4)
Fasting glucose (mmol/L)	5.5 (5.2–6)	5.5 (5.2–6)	>0.05	5.5 (5.2–6)
Insulin, (μU/mL)	8.1 (5.5–12.1)	8.2 (5.5–12.6)	>0.05	8.1 (5.5–12.2)
Haemoglobin A1c (g/dL)	0.5 (0.4–0.5)	0.5 (0.4–0.5)	>0.05	0.5 (0.4–0.5)
C-reactive protein (mg/L)	1.8 (0.9–3.7)	2 (1.1–3.9)	9.50E−05	1.8 (1–3.8)
Electrocardiogram				
PR interval (ms)	166 (152–184)	170 (152–192)	4.10E−04	168 (152–186)
QRS interval (ms)	90 (84–100)	92 (84–102)	4.00E−02	92 (84–100)
QT interval (ms)	406 (382–428)	412 (390–438)	1.70E−09	406 (384–430)
Prevalent comorbidities				
Diabetes	402 (11.2%)	149 (12.7%)	>0.05	551 (11.6%)
Chronic kidney disease (eGFR < 60)	1332 (37.1%)	465 (39.7%)	>0.05	1797 (37.7%)
Hypertension	1827 (50.8%)	676 (57.7%)	5.50E−05	2503 (52.5%)
Coronary artery disease	700 (19.5%)	278 (23.7%)	2.10E−03	978 (20.5%)
Myocardial infarction	380 (10.6%)	161 (13.7%)	3.60E−03	541 (11.4%)
Heart failure	47 (1.3%)	30 (2.6%)	4.80E−03	77 (1.6%)
Medication				
Antihypertensive medication use	2124 (59.1%)	814 (69.5%)	3.20E−10	2938 (61.7%)
ACE inhibitors	423 (11.8%)	188 (16%)	1.80E−04	611 (12.8%)
Angiotensin II receptor blockers	451 (12.6%)	198 (16.9%)	2.00E−04	649 (13.6%)
Beta-blocking agents	1137 (31.6%)	459 (39.2%)	2.60E−06	1596 (33.5%)
Calcium channel blockers	529 (14.7%)	213 (18.2%)	5.40E−03	742 (15.6%)
Diuretics	1013 (28.2%)	407 (34.7%)	2.60E−05	1420 (29.8%)
Other	20 (0.6%)	7 (0.6%)	>0.05	27 (0.6%)
Antiarrythmic medication use	15 (0.4%)	23 (2%)	6.50E−07	38 (0.8%)
Antiarrhythmics, class Ia/Ib	3 (0.1%)	11 (0.9%)	1.20E−05	14 (0.3%)
Antiarrhythmics, class III	12 (0.3%)	12 (1%)	7.80E−03	24 (0.5%)
Anticoagulation medication use	556 (15.5%)	204 (17.4%)	>0.05	760 (15.9%)

Continued

Table 1 *Continued*

Variable	Non-case	Incident AF	*P*	Total
Diabetes medication use	206 (5.7%)	67 (5.7%)	>0.05	273 (5.7%)
Statin use	752 (20.9%)	279 (23.8%)	4.20E−02	1031 (21.6%)

Numbers represent the median (interquartile range) and count (%) for continuous and categorical variables, respectively. *P* values for continuous variables were determined with the Wilcoxon signed-rank test and the *F*-test for categorical variables. All reported *P* values are two-sided.
AF, atrial fibrillation; BMI body mass index; eGFR, estimated glomerular filtration rate; HDL cholesterol, high-density lipoprotein cholesterol; LDL cholesterol, low-density lipoprotein cholesterol; ACE inhibitors, angiotensin-converting enzyme inhibitors.

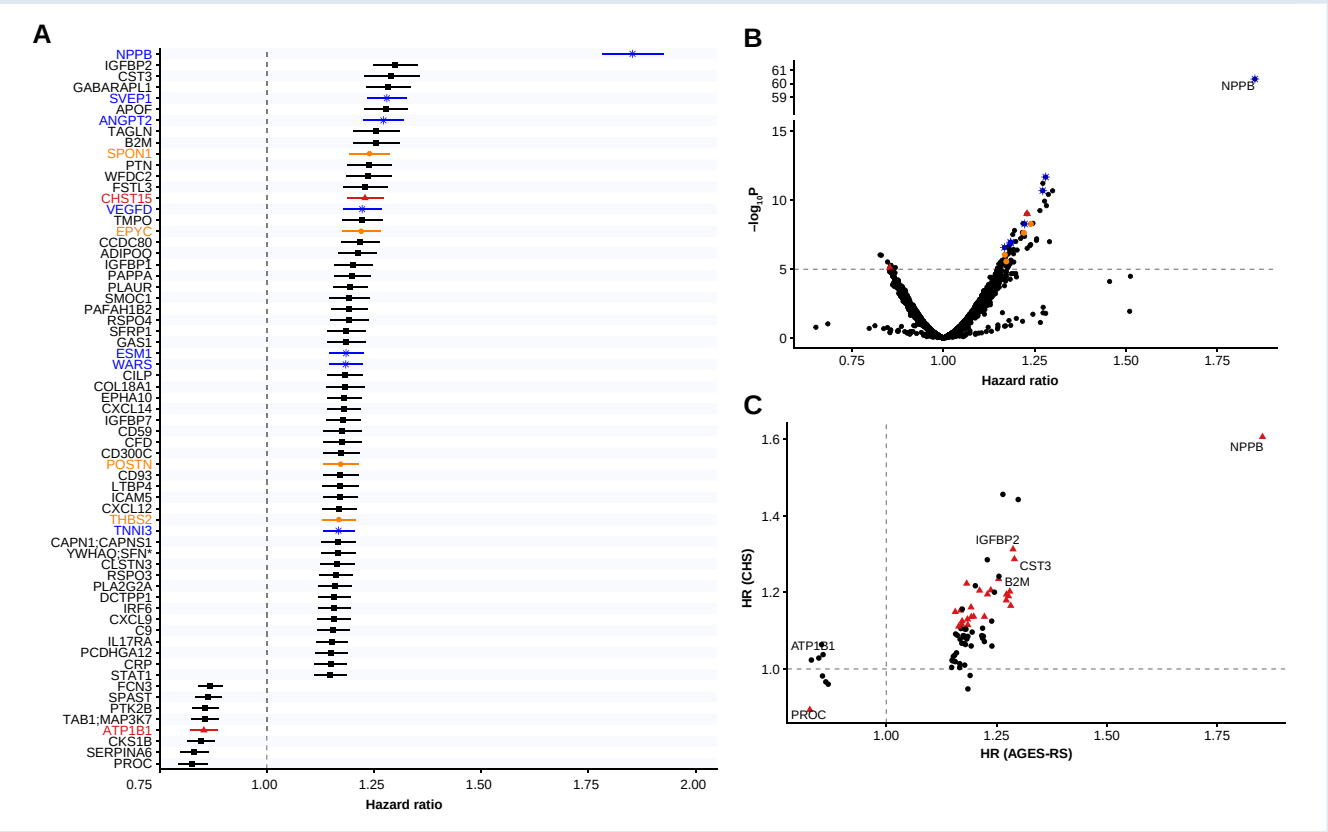


Figure 2 (*A*) Study-wide significant proteins associated with incident AF in Model 1. Forest plot of HRs where error bars represent SE. (*B*) Volcano plot of all protein HRs. The broken line corresponds to the study-wide significant Bonferroni threshold. Highlighted proteins belong to the pathways angiogenesis (blue star), extracellular matrix (orange circle), and cardiac remodelling or conductivity (red triangle). The strongest association was for NPPB (NT-proBNP; HR = 1.85, $P = 4.65 \times 10^{-61}$). (*C*) The results of the replication analysis performed by CHS. In total, 46/65 (71%) protein associations were found to be statistically significant ($P < 0.05$) in CHS (black circle), of which 24 (37%) passed the replication threshold ($P < 0.05/74$, red triangle). There was complete agreement between the HRs of replicated proteins. Natriuretic peptide B was observed to have the strongest significant association with incident AF in both studies. The HR coefficients are shown per SD of the protein serum levels for both AGES-RS and CHS.

be subjected to a sensitivity analysis. The causal analysis suggested a protective effect of genetically determined TAGLN levels for AF (*Figure 4A*, Supplementary material online, *Table S3*). This contradicted the findings of the observational study, where measured serum levels of TAGLN were associated with increased risk of AF (*Figure 4A*, Supplementary material online, *Table S3*). Importantly, however, the observational association of TAGLN with AF was NPPB-dependent, considering its attenuation in model 2. CHST15 had directionally consistent observational and causal estimates (*Figure 4A*), with increased measured levels and genetically determined levels associated with increased risk of AF.

Several proteins were identified in recent MR studies[13,14] as having a potential causal role in AF, independently of observational data. As unknown confounders can mask the presence of putative causal protein associations in an observational analysis, an extended forward MR analysis was performed for all 1294 proteins with available cis-acting instruments from our data (see Supplementary material online, *Table S9*). Six proteins (CHMP3, COL15A1, DUSP13, MANBA, QSOX2, and SRL)

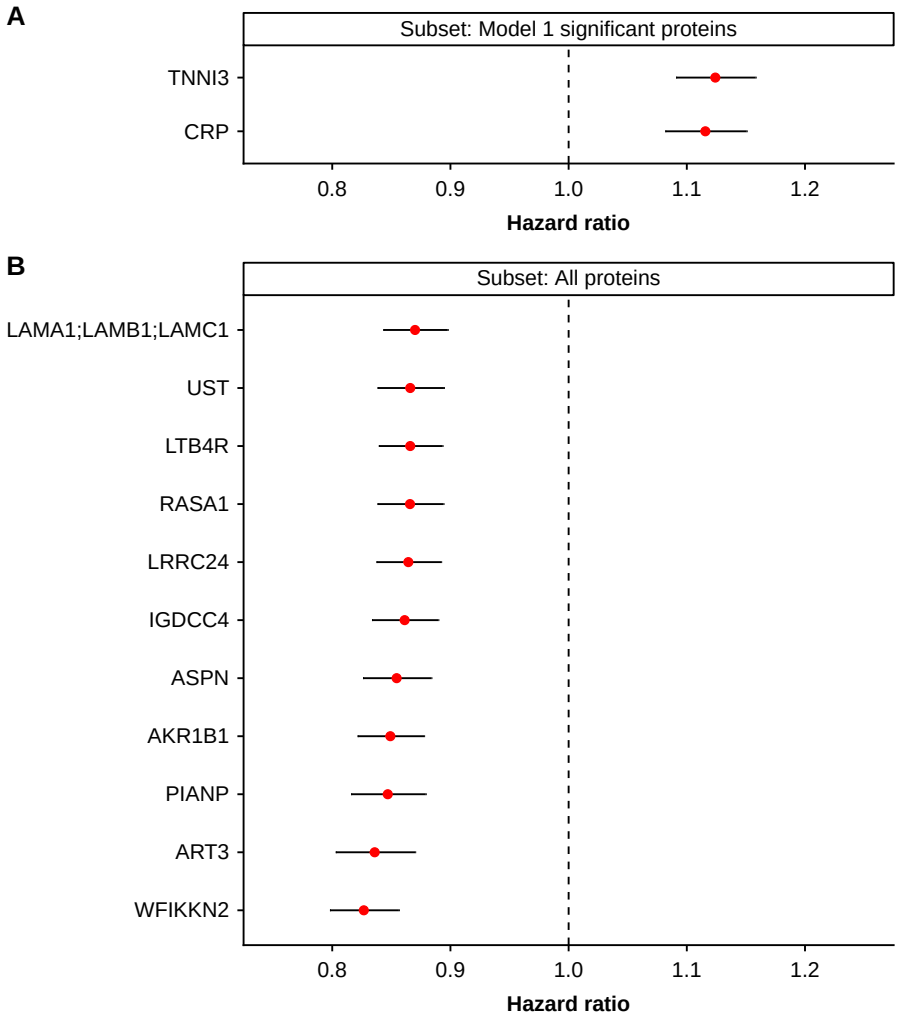


Figure 3 Forest plot of study-wide significant, NPPB-independent (Model 2) protein associations with incident AF in AGES-RS. (*A*) Of the 65 proteins identified with Model 1, TNNI3 and CRP were reaffirmed as associated with incident AF ($P < 0.05/4781$) in AGES-RS. (*B*) When the analysis included the entire AGES-RS protein panel, 11 new proteins were identified ($P < 0.05/4781$), all of which were protective for incident AF.

had significant (FDR < 0.05) support for causality (*Figure 4C*, Supplementary material online, *Figures S5–S10* and *Table S10*). Of those, COL15A1, DUSP13, MANBA, and SRL had more than two instruments and could be subjected to the sensitivity analysis. MANBA had a significant intercept term ($P < 0.05$) in the MR-Egger analysis suggesting the presence of horizontal pleiotropy, whereas COL15A1, DUSP13, and SRL passed both phases. When the causal estimates were compared with their respective observational counterparts, CHMP3 was significant ($P < 0.05/6$) but had an inconsistent observational estimate (*Figure 4C*). Of the proteins with causal support in the MR analysis, COL15A1, MANBA, and QSOX2 have been listed as druggable targets.[36]

Serum protein levels affected by genetic liability to atrial fibrillation

A reverse MR analysis was performed to evaluate if the observed protein changes prior to disease onset were likely to arise as a consequence of the genetic liability of AF, as determined by the collective effect of common genetic risk variants.[22] Of the 76 proteins identified in the observational analysis, 31 were study-wide significant (FDR < 0.05) in the reverse MR analysis (see Supplementary material online, *Table S3*). These included proteins such as CHST15 and NPPB, whereas the only protein identified by Model 2 was RASA1. Of these 31 proteins, 13 unique proteins passed the sensitivity analysis (see Methods). Although most proteins passed Cochran's *Q* test, NPPB displayed evidence of instrument heterogeneity (see Supplementary material online, *Table S3*). A leave-one-out locus analysis did not suggest that any single locus was driving the causal signal of AF (see Supplementary material online, *Table S11*) although the removal of a single locus on chromosome 4 affected the significance of the reverse causal estimate for APOF (see Supplementary material online, *Table S11*). For completeness, the reverse MR analysis was also performed on the entire AGES-RS protein panel. In total, 322 proteins were study-wide significant in this analysis (FDR < 0.05) (see Supplementary material online, *Table S10*). It is of note that the causal estimates from the reverse MR analysis were in near complete agreement with their respective observational counterparts (*Figure 4B and D*, Supplementary material online, *Tables S3* and *S11*).

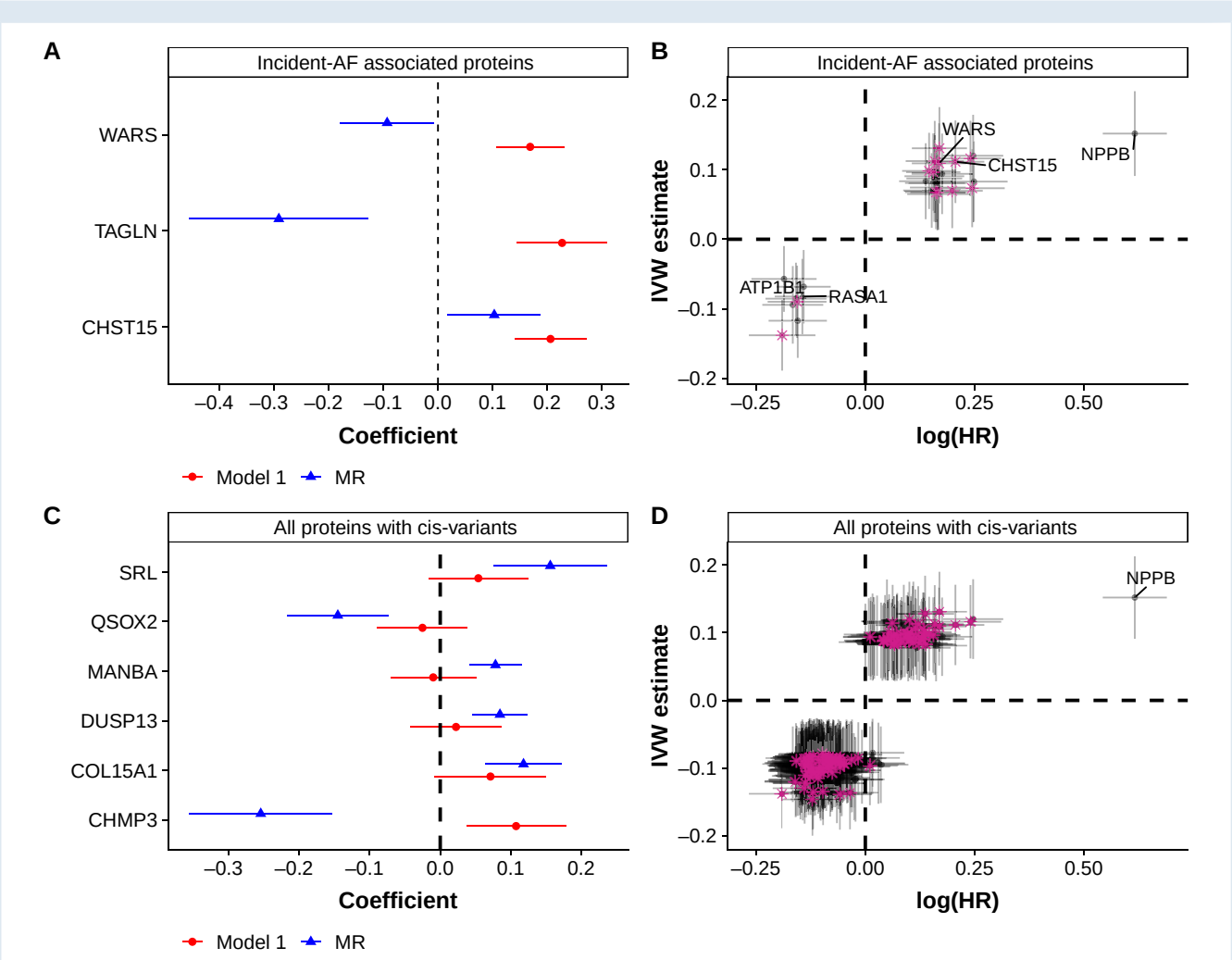


Figure 4 The bidirectional, two-sample MR analysis was performed for two sets of proteins: incident AF-associated proteins (*A*, *B*) and all proteins with cis-variants (*C*, *D*). (*A*, *C*) Forest plot comparison of log HRs from the observational analysis (red circles) and estimates from the forward MR analysis (blue triangles), shown for proteins that were nominally significant ($P < 0.05$) in the forward MR analysis. (*B*, *D*) Scatter plot comparison of log HRs from the observational analysis and estimates from the reverse MR analysis, shown for all proteins significant (FDR < 0.05) in the reverse MR analysis. Pink stars represent causal estimates which passed all MR sensitivity tests. (*A*) Three proteins, identified in Model 1 (CHST15, WARS, and TAGLN), were nominally significant ($P < 0.05$) in the forward MR analysis. CHST15 had consistent observational and causal estimates. (*B*) Thirty-one proteins were significant (FDR < 0.05) in the reverse MR analysis, of which 19 passed the sensitivity testing. All causal coefficients were consistent with their respective observational estimates. (*C*) Six proteins were significant (FDR < 0.05) in the forward MR analysis when all proteins, irrespective of their observational significance, were considered. Three proteins, SRL, COL15A1, and DUSP13, had consistent causal and observational estimates. (*D*) Three hundred eleven proteins were significant (FDR < 0.05) in the reverse MR analysis of which 100 passed the sensitivity testing. Almost all causal estimates were directionally consistent with their observational counterparts.

Polygenic risk score as an expanded measure of the genetic liability to atrial fibrillation

The reverse MR analysis utilized a highly specific subset of independent genome-wide significant genetic variants for AF. A genome-wide AF-PRS based on a less stringent SNP inclusion can be considered as an expanded assessment of the genetic liability of AF criterion. Therefore, it was of interest to see how such a measure would affect the AF protein profile. The primary AF-PRS constructed for AGES-RS ($r^2 < 0.2$, $P < 0.5$, #SNPs = 297 997; Supplementary material online, *Table S12*) was significantly higher in the incident AF group (10-year follow-up, $P = 3.42 \times 10^{-168}$), and an exponential increase was seen in the percentage of AF incident cases for successive 2%-PR-score-percentile bands (*Figure 5A and B*). The AF-PRS was significantly associated with incident AF (HR = 2.54; $P = 5.92 \times 10^{-254}$), after age and sex adjustment, and had a *c*-statistic of 0.79 [standard error (SE) = 0.006] for the full follow-up time (see Supplementary material online, *Table S12*).

The AF-PRS was significantly (FDR < 0.05) associated with the serum levels of 67 of the 76 proteins associated with incident AF in the observational analysis (*Figure 5C*, Supplementary material online, *Table S13*). These included top markers for incident AF such as NPPB, ANGPT2, and TAGLN (*Figure 6*, Supplementary material online, *Table S14*), as well as Model 2–specific proteins like LAMA1/LAMB1/LAMC1 and ASPN. Furthermore, almost all (27/31) of the reverse MR–significant proteins were also associated with the AF-PRS

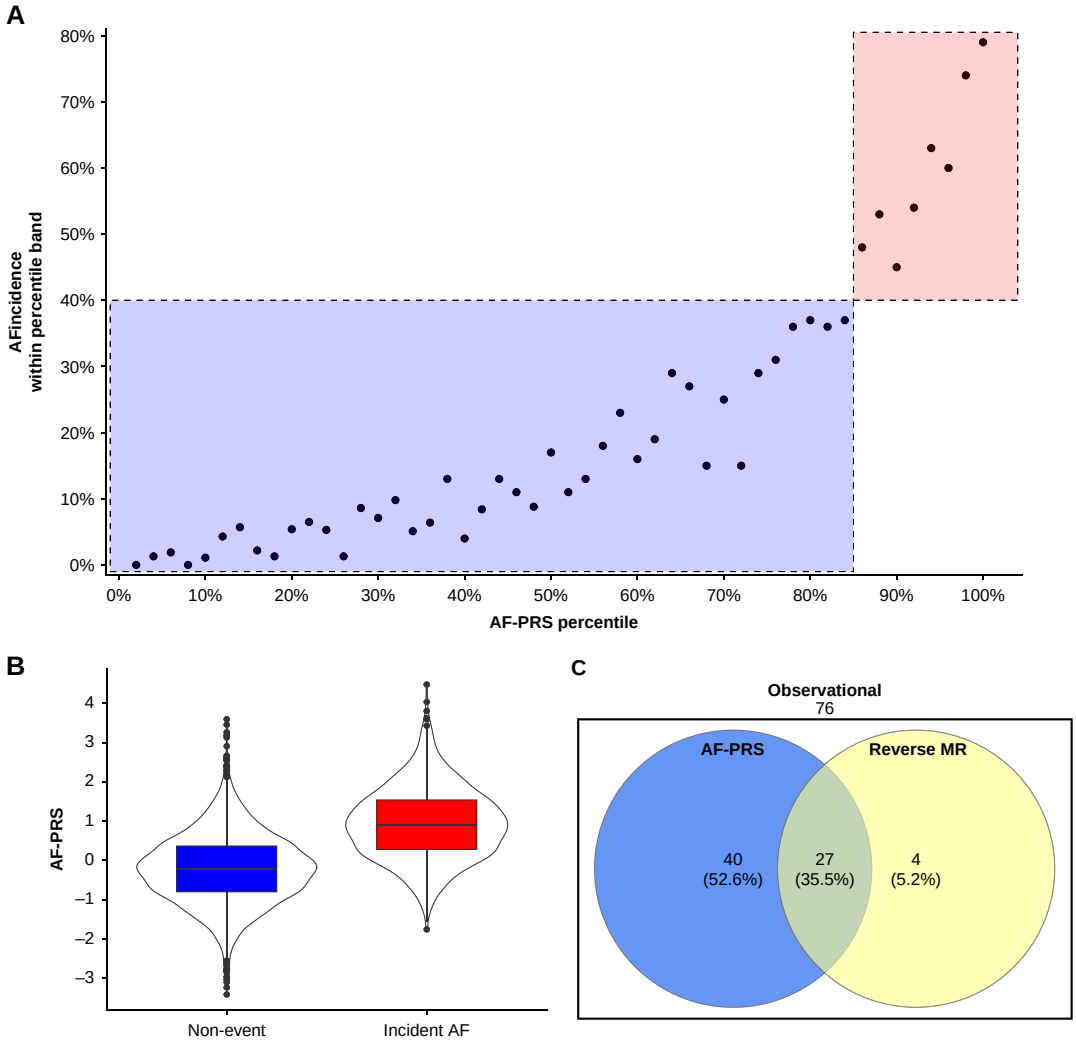


Figure 5 (*A*) Percentage of incident AF cases per successive 2% AF-PRS percentile bands based on 10-year survival. Each point represents ~92 (4601/50) individuals. A good separation between low- and high-risk groups is observed as the ratio of events increases with each subsequent percentile band. For example, the event ratio increases ranges from <5 to 40% between the 2nd and 85th percentiles (blue lower rectangle) but then increases rapidly up to 80% for the remaining participants (red upper rectangle). (*B*) The distribution of the AF-PRS between the AF incident and non-event groups based on a 10-year follow-up. The mean standardized AF-PRS was −0.21 (SD: 0.89) and 0.94 (SD: 0.93) for the non-event and incident AF groups, respectively (*t*-test, $P = 3.42 \times 10^{-168}$). (*C*) Overlap between incident AF-associated proteins from the observational analysis, significant (FDR < 0.05) proteins from the reverse MR analysis and proteins significantly (FDR < 0.05) associated with the AF-PRS. The overlap for the AF-PRS analysis was much greater compared with the reverse MR or 88.2 vs. 40.8%, respectively.

(*Figure 5C*). Proteins affected by the AF-PRS in individuals without prevalent AF may highlight causal processes that may be dysregulated at early stages of the condition. Extending the analysis to the entire protein panel, we found the AF-PRS to be significantly associated (FDR < 0.05) with 257 and 72 unique proteins in Models 1 and 2, respectively, with only 45 overlapping. A GSEA showed that the AF-PRS was associated with up-regulation of proteins involved in RNA binding, response to cytokine stimulus, and collagen-containing extracellular matrix (see Supplementary material online, *Table S14*). In contrast, in Model 2 adjusting for NPPB, the AF-PRS was particularly strongly associated with the down-regulation of proteins involved in axonogenesis and neuron projection but also the up-regulation of DNA binding, cardiac muscle cell apoptotic process, and lymphocyte-mediated immunity (see Supplementary material online, *Table S14*).

Predictive performance of the proteomic signature and atrial fibrillation-polygenic risk score

Finally, we aimed to assess the predictive value of serum protein levels and the AF-PRS over clinical risk factors in the AGES-RS cohort. Four proteins were selected in over 70% of all 2000 bootstrap iterations: NPPB (100%), TNNI3 (94%), CRP (81%), and SPAST (72%) (see Supplementary material online, *Table S15*). Although the 10-year AUC_t statistic increased when all four proteins were added to Model 1, NPPB was the main contributor to the observed increase (see Supplementary material online, *Figure S11*). Adding a penalized spline basis for NPPB did not affect the AUC_t statistic (see Supplementary material online, *Table S8*).

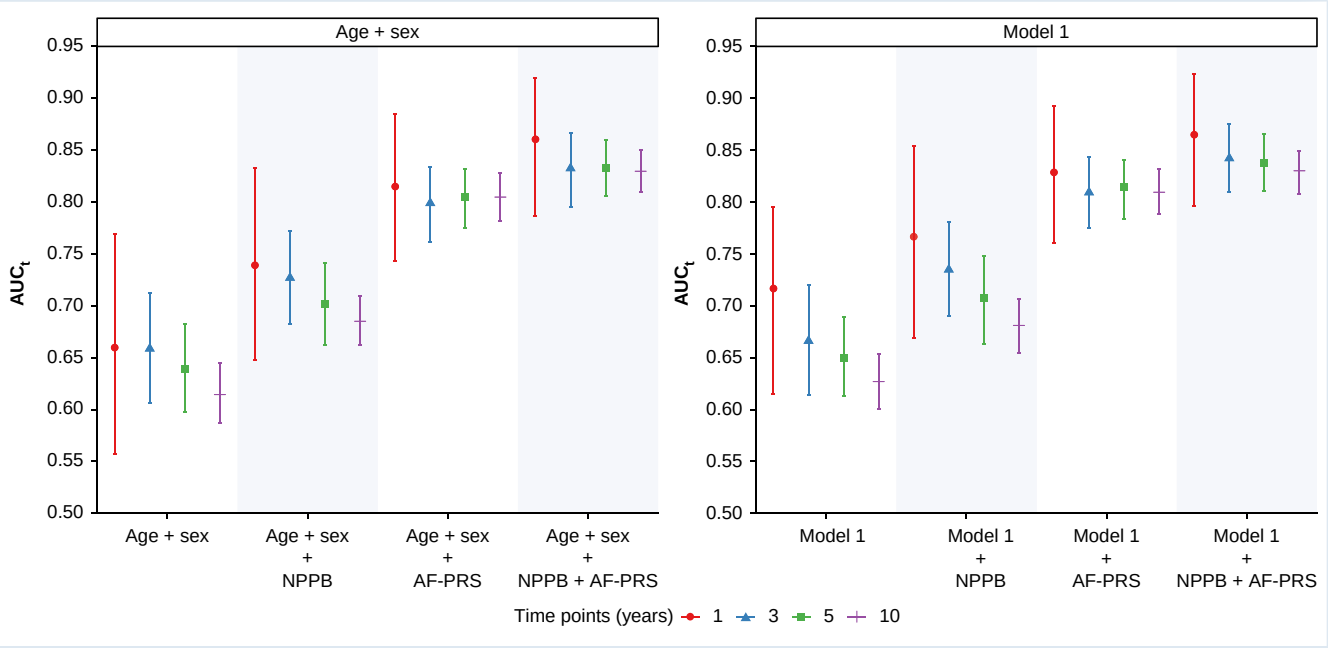


Figure 6 Cumulative/dynamic time-dependent ROC curve estimation of the time-dependent AUC-statistic (AUC_t) evaluated at 1, 3, 5, and 10 years. Error bars represent bootstrap percentile intervals (2.5th and 97.5th percentiles), constructed on the data left out of the bootstrap sample at each iteration. The covariates of Model 1 were CHARGE-AF risk factors, sex, and eGFR. The greatest variation in AUC_t was observed for events occurring within 1 year. Adding the CHARGE-AF risk factors netted little gain in AUC_t in comparison with adding NPPB or the AF-PRS, irrespective of the time point under consideration. The effect of NPPB diminished over time, whereas all models including the AF-PRS were relatively stable. The largest AUC_t was seen for models incorporating both NPPB and the AF-PRS ($AUC_t = 0.83$ at 10 years).

For events occurring within 10 years, clinical risk factors provided little increase in predictive value [$AUC_t = 0.64$ (0.60; 0.65)] compared with age and sex alone [$AUC_t = 0.62$ (0.59; 0.65)] (*Figure 6*, Supplementary material online, *Table S16*). In comparison, adding NPPB levels to the base model of age and sex yielded an AUC_t of 0.69 (0.66; 0.71), while adding AF-PRS as a covariate netted the biggest differential in the *c*-statistic [$AUC_t = 0.81$ (0.78; 0.83)]. The best prediction was provided by combining NPPB and the AF-PRS with age and sex [$AUC_t = 0.84$ (0.81; 0.85)], suggesting that complementary information is captured in these two variables. There was no difference between models with or without clinical risk factors when both NPPB and the AF-PRS were included [$AUC_t = 0.84$ (0.81; 0.85) for both], demonstrating that clinical information did not add much information over these two predictors.

The difference in the time dynamics of the models was also noteworthy. The base model of age and sex had a clear time-dependent pattern, with better performance observed for events occurring closer to baseline (*Figure 6*, Supplementary material online, *Figure S12*). This pattern extended to both NPPB and clinical risk factors, indicating that these factors provide better predictive value for events closer to baseline (*Figure 6*, Supplementary material online, *Figure S12A and B*). Conversely, the AF-PRS had a relatively stable AUC_t across all cumulative time points and thus provided a greater increase in predictive performance at later time points where other variables were lacking (*Figure 6*, Supplementary material online, *Figure S12C and D* and *Table S16*).

Discussion

This study joined observational and genetic components of serum protein associations occurring prior to clinical onset of AF to provide a unique holistic view of the disease. Furthermore, a comprehensive comparison of clinical, genetic, and protein risk factors in a population-based cohort has been provided, which has hitherto been unavailable.

The proteomic signature of incident AF in AGES-RS consisted of 76 unique proteins (corresponding to 88 aptamers), of which 63 (83%) were novel and 29 [38% (Model 1, 24/65; Model 2, 5/11)] were replicated successfully in CHS. This signature included both NPPB-dependent and NPPB-independent components, as well as putative causal candidates. We found NPPB to be the only protein that provided significant improvement in the identification of individuals at risk of AF over clinical risk factors, while the best prediction was obtained when NPPB was combined with genetic information that especially improved long-term prediction of AF.

Ion channel dysfunction and structural remodelling are central to the pathophysiology of AF.[1] Among the novel associations identified in this study were proteins relevant to these processes, which may point to molecular markers of a vulnerable state preceding AF diagnosis. ATP1B1[37] sustains the Na/K gradient, and genetic association has been established between ATP1B1 and the QT complex.[38] ASPN is an extracellular matrix protein which can be released by cardiac fibroblasts as an effort to remodel the heart after a traumatic event such as heart failure.[39] The laminin complex LAMA1/LAMB1/LAMC1 has been linked to a dysfunction of cardiac conductivity,[40] along with differential expressions between individuals with and without AF.[41] All three were protective for incident AF in AGES-RS, and the reverse MR analysis revealed that the genetic risk of AF was significantly associated with a down-regulation of ATP1B1 serum levels.

We identified several biologically relevant proteins with a support for a causal role in AF. CHST15 was consistently associated with increased risk of incident AF in both the causal and observational analyses. Although the mechanism underlying the adverse effects of elevated CHST15 levels remain unidentified, it is known that the protein participates in the biosynthesis of chondroitin sulfate E,[42] a polymer which can drive post-MI

sympathetic denervation in cardiac scars and consequently cause arrhythmias.[43] TAGLN, which is expressed in smooth muscle,[44] was found to be protective for AF in the causal analysis, contradicting the observational results. This might be due to pleiotropic effects, which could not be tested as TAGLN had a single cis-acting variant or because the observational effect mainly reflects an NPPB-dependent association with AF as this contradiction disappeared when adjusting for NPPB serum levels (see Supplementary material online, *Figure S12*). SRL is a calcium-binding protein located in the longitudinal sarcoplasmic reticulum in skeletal and cardiac muscles that may play a role in maintaining cardiac function under stress.[45] COL15A1 is necessary for extracellular matrix organization in the heart, and its deficiency predisposes to cardiomyopathy.[46] The results of these studies suggested a protective function with respect to cardiac health, yet the causal analysis implied that elevated serum levels of either protein increased the risk of AF. Finally, DUSP13 was found to be risk-increasing for AF and has been discussed in recent AF-MR studies[13,14] due to its strong evidence of colocalization between the genetic signals of the protein and AF and role in cardiac hypertrophy and failure.[47]

Natriuretic peptide B (NPPB) (generally referred to as NT-proBNP) is a well-established predictor of AF[48–51] and a dependable biomarker for cardiac health,[52] as its levels are increased in response to cardiac distress. We found that TINNI3 (cardiac troponin I) and CRP were associated with AF independently of NPPB. Both proteins have been previously associated with AF, either directly[53] or indirectly through conditions such as cardiomyopathy[54,55] and coronary heart disease.[56,57] The independent association of CRP may give support for atrial inflammatory signalling, an emerging model of AF pathophysiology.[7] The overall attenuation of other protein associations after adjusting for NPPB suggest that the proteomic profile for incident AF may largely reflect the effects of NPPB on the circulating proteome or an underlying process which elevates serum levels of NPPB. Indeed, the reverse causation and AF-PRS analyses showed that NPPB levels are affected by the genetic liability to AF, together with a large proportion of the AF-associated proteins. As NPPB itself is unlikely to be a causal factor for AF, as has been demonstrated in previous[58] and the current MR analysis, this further suggests that NPPB levels change downstream of the causal processes that are captured by the genetic risk factors for AF. The proteins affected by the reverse MR and AF-PRS may provide insight into such causal processes. For example, after adjusting for NPPB levels, the AF-PRS showed a clear association with the down-regulation of proteins involved in axonogenesis, hinting at a neuronal component to the development of AF. These included neural cell adhesion molecule 1 (NCAM1), which in the Human Protein Atlas[59] is expressed in a cluster of genes involved in heart contraction. Given our data, it is difficult to ascertain the causes of neuronal protein down-regulation in the circulation, and further studies are required to evaluate if and how they relate to the cardiac conduction system.

A discrepancy between the observational and causal estimates was observed for over half (5/9) of the proteins, a phenomenon which has been reported in other studies.[19,60] This discordance implies that the genetically determined levels of circulating proteins, which are the components evaluated in an MR analysis, often differ from the remaining variance in measured protein levels which might be strongly confounded by external factors. An example of such confounding would be the process which elevates NPPB serum levels, given the attenuation of most proteins association after adjusting for NPPB. Thus, the observational associations may not always be able to detect true causal relationships and a discrepancy between causal and observational estimates does not rule out the possibility of true causality. At the same time, MR has several drawbacks[61] and we cannot ignore the possibility of false positive results. Furthermore, we cannot rule out the possibility of molecular pleiotropy where some of the SNPs selected as genetic instruments for serum protein levels may also regulate the corresponding, or other, gene or protein expression in different tissues, which may represent the true causal pathway.

Finally, despite the large number of replicated AF-associated proteins across studies, NPPB was the only protein which improved the classification capabilities of the baseline risk model. The inclusion of NPPB had the largest effect when the cumulative time period under consideration was relatively short, up to 3 years. An even greater increase, which was consistently high over time, was observed for a model including the AF-PRS without any proteomic variables. However, the greatest discriminating ability was realized when both variables were used. This result emphasizes the clinical relevance of NPPB with respect to AF, as measuring this single protein might be sufficient to assess risk of AF developing within a few years. Adding PRSs to established clinical risk models is a growing trend in cardiometabolic research, with reported AUC/*c*-statistic values ranging from 0.61 to 0.82.[12,62] Our study also demonstrates the potential of PRSs as predictors for disease incidence, although their feasibility to clinical application remains an open question.[62] The differential separation achieved by AF-PRSs across studies might be in part due to differences in recruitment strategy, age of cohort at baseline, or the size of the base GWAS providing the estimates used to construct the AF-PRS. In our study, we additionally cannot exclude a potential influence of unknown sample overlap between the AGES-RS cohort and the AF GWAS.

Our study includes several limitations. As AGES-RS is a prospective population-based cohort reliant on ECGs and hospital records, the true incidence and prevalence of AF may have been underestimated due to silent and paroxysmal AF.[1] Furthermore, direct data on important risk factors such left atrial size, left ventricular function, and intermediate atrial phenotypes[1] were not measured. Despite these limitations, proteins associated with incident AF were identified in this study and replicated and validated in external cohorts, providing candidates for future clinical and functional studies.

In conclusion, we demonstrate a shift in the serum proteome that is associated with the risk of incident AF in a population-based cohort. This shift seems to reflect an NPPB-dependent response to the genetic liability of AF. Causal analysis shows that the observational associations for incident AF largely reflect the said genetic liability but prior to AF incidence. Importantly, we still identify a number of novel causal candidates for future functional studies. Finally, we show that NPPB and AF-PRS form potent indicators for AF incidence, both for short-term and long-term prediction.

Translational perspective

The study identifies multiple proteins associated with incident atrial fibrillation (AF) in the Age, Gene/Environment Susceptibility-Reykjavik study (AGES-RS), a few of which were independent of N-terminal prohormone of brain natriuretic peptide (NT-proBNP). Seven putative causal candidate proteins were identified, some with potential therapeutic relevance. However, the causal analysis suggested that most of the incident AF protein signature is a response to the genetic liability of AF. N-terminal prohormone of brain natriuretic peptide serum levels achieved greater discrimination for incident AF cases compared with risk models consisting solely of clinical risk factors, especially for events occurring at earlier stages of follow-up. A polygenic risk score for AF achieved a good separation between high- and low-risk participants in AGES-RS and further improved the predictive performance of NT-proBNP, although the two complement each other.

Supplementary material

Supplementary material is available at *Europace* online.

Acknowledgements

The authors acknowledge the contribution of the Icelandic Heart Association (IHA) staff to AGES-Reykjavik, as well as the involvement of all study

participants. The graphical abstract was created with BioRender.com (https://www.BioRender.com).

Funding

AGES-RS: The National Institute of Aging contracts N01-AG-12100 and HHSN271201200022C for Vi.G. financed the study. Vi.G. received funding from the National Institute of Aging (1R01AG065596-01A1), and I.H.A received a grant from Althingi (the Icelandic Parliament). T.J. and Va.G. were supported by the Icelandic Centre for Research (grant no. 206692-051) and the University of Iceland Research Fund. The protein measurements in AGES-RS were supported by Novartis Biomedical Research. CHS: This research was supported by contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and 75N92021D00006 and grants U01HL080295, U01HL130114, and HL144483 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the NIA. J.S.F was supported by National Heart Lung Blood Institute grant R01HL149706. A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org.

Conflict of interest: L.J. is an employee and stockholder of Novartis. B.M.P. serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. All other authors have no conflict of interests to declare.

Data availability

The custom-design Novartis SOMAscan is available through a collaboration agreement with the Novartis Institutes for Biomedical Research (lori.jennings@novartis.com). Data from the AGES-Reykjavik study are available through collaboration (AGES_data_request@hjarta.is) under a data usage agreement with the IHA. Data from the CHS study are available through a collaboration agreement. All access to data is controlled via the use of a subject-signed informed consent authorization. All other data supporting the conclusions of the paper are presented in the main text and freely available as a supplement to this manuscript.

References

1. Andrade J, Khairy P, Dobrev D, Nattel S. The clinical profile and pathophysiology of atrial fibrillation. *Circ Res* 2014;**114**:1453–68.
2. Ringborg A, Nieuwlaat R, Lindgren P, Jönsson B, Fidan D, Maggioni AP *et al.* Costs of atrial fibrillation in five European countries: results from the Euro Heart Survey on atrial fibrillation. *Europace* 2008;**10**:403–11.
3. Kim MH, Johnston SS, Chu B-C, Dalal MR, Schulman KL. Estimation of total incremental health care costs in patients with atrial fibrillation in the United States. *Circ Cardiovasc Qual Outcomes* 2011;**4**:313–20.
4. Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: an increasing epidemic and public health challenge. *Int J Stroke* 2021;**16**:217–21.
5. Nattel S, Heijman J, Zhou L, Dobrev D. Molecular basis of atrial fibrillation pathophysiology and therapy: a translational perspective. *Circ Res* 2020;**127**:51–72.
6. Roselli C, Rienstra M, Ellinor PT. Genetics of atrial fibrillation in 2020. *Circ Res* 2020;**127**:21–33.
7. Dobrev D, Heijman J, Hiram R, Li N, Nattel S. Inflammatory signalling in atrial cardiomyocytes: a novel unifying principle in atrial fibrillation pathophysiology. *Nat Rev Cardiol* 2023;**20**:145–67.
8. Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB *et al.* Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc* 2013;**2**:e000102.
9. Huxley RR, Lopez FL, Folsom AR, Agarwal SK, Loehr LR, Soliman EZ *et al.* Absolute and attributable risks of atrial fibrillation in relation to optimal and borderline risk factors. *Circulation* 2011;**123**:1501–8.
10. Norby FL, Tang W, Pankow JS, Lutsey PL, Alonso A, Steffan B *et al.* Proteomics and risk of atrial fibrillation in older adults (from the Atherosclerosis Risk in Communities [ARIC] study). *Am J Cardiol* 2021;**161**:42–50.
11. Ko D, Benson MD, Ngo D, Yang Q, Larson MG, Wang TJ *et al.* Proteomics profiling and risk of new-onset atrial fibrillation: Framingham Heart Study. *J Am Heart Assoc* 2019;**8**:e010976.
12. Marston NA, Garfinkel AC, Kamanu FK, Melloni GM, Roselli C, Jarolim P *et al.* A polygenic risk score predicts atrial fibrillation in cardiovascular disease. *Eur Heart J* 2023;**44**:221–31.
13. Wang Q, Richardson TG, Sanderson E, Tudball MJ, Ala-Korpela M, Davey Smith G *et al.* A phenome-wide bidirectional Mendelian randomization analysis of atrial fibrillation. *Int J Epidemiol* 2022;**51**:1153–66.
14. Ning Z, Huang Y, Lu H, Zhou Y, Tu T, Ouyang F, *et al.* Novel drug targets for atrial fibrillation identified through Mendelian randomization analysis of the blood proteome. *Cardiovasc Drugs Ther* 2023. doi:10.1007/s10557-023-07467-8 (Online ahead of print).
15. Harris TB, Launer LJ, Eiriksdottir G, Kjartansson O, Jonsson PV, Sigurdsson G *et al.* Age, Gene/Environment Susceptibility–Reykjavik study: multidisciplinary applied phenomics. *Am J Epidemiol* 2007;**165**:1076–87.
16. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA *et al.* The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* 1991;**1**:263–76.
17. Smith GD, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;**23**:R89–98.
18. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science* 2018;**361**:769–73.
19. Gudmundsdottir V, Zaghlool SB, Emilsson V, Aspelund T, Ilkov M, Gudmundsson EF *et al.* Circulating protein signatures and causal candidates for type 2 diabetes. *Diabetes* 2020;**69**:1843–53.
20. Gudjonsson A, Gudmundsdottir V, Axelsson GT, Gudmundsson EF, Jonsson BG, Launer LJ *et al.* A genome-wide association study of serum proteins reveals shared loci with common diseases. *Nat Commun* 2022;**13**:480.
21. Therneau TM. A Package for Survival Analysis in R. 2022. Available from https://CRAN.R-project.org/package=survival (1 January 2022, date last accessed).
22. Nielsen JB, Thorolfsdottir RB, Fritsche LG, Zhou W, Skov MW, Graham SE *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet* 2018;**50**:1234–9.
23. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* 2017;**32**:377–89.
24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–75.
25. Gkatzionis A, Burgess S, Newcombe PJ. Statistical methods for cis-Mendelian randomization with two-sample summary-level data. *Genet Epidemiol* 2023;**47**:3–25.
26. Hemani G, Zheng J, Elsworth B, Wade K, Baird D, Haberland V *et al.* The MR-base platform supports systematic causal inference across the human phenome. *eLife* 2018;**7**:e34408.
27. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med* 2016;**35**:1880–906.
28. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur J Epidemiol* 2018;**33**:947–52.
29. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016;**40**:304–14.
30. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018;**50**:1219–24.
31. Korotkevich G, Sukhov V, Budin N, Sergushichev A, *et al.* Fast gene set enrichment analysis. *bioRxiv* 2021. https://doi.org/10.1101/060012
32. Blanche P, Dartigues J-F, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* 2013;**32**:5381–97.
33. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Med Res Methodol* 2017;**17**:53.
34. Heagerty PJ, Saha-Chaudhuri P. survivalROC: Time-Dependent ROC Curve Estimation from Censored Survival Data. 2022. Available from: https://CRAN.R-project.org/package=survivalROC (1 January 2022, date last accessed).
35. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000;**56**:337–44.
36. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J *et al.* The druggable genome and support for target identification and validation in drug development. *Sci Transl Med* 2017;**9**:eaag1166.
37. Barwe SP, Jordan MC, Skay A, Inge L, Rajasekaran SA, Wolle D *et al.* Dysfunction of ouabain-induced cardiac contractility in mice with heart-specific ablation of Na, K-ATPase beta1-subunit. *J Mol Cell Cardiol* 2009;**47**:552–60.
38. Pfeufer A, Sanna S, Arking DE, Müller M, Gateva V, Fuchsberger C *et al.* Common variants at ten loci modulate the QT interval duration in the QTSCD study. *Nat Genet* 2009;**41**:407–14.
39. Huang C, Sharma A, Thakur R, Rai D, Katiki M, Germano JF *et al.* Asporin, an extracellular matrix protein, is a beneficial regulator of cardiac remodeling. *Matrix Biol* 2022;**110**:40–59.
40. Malan D, Reppel M, Dobrowolski R, Roell W, Smyth N, Hescheler J *et al.* Lack of laminin γ1 in embryonic stem cell-derived cardiomyocytes causes inhomogeneous electrical spreading despite intact differentiation and function. *Stem Cells* 2009;**27**:88–99.

41. Liu B, Li X, Zhao C, Wang Y, Lv M, Shi X *et al.* Proteomic analysis of atrial appendages revealed the pathophysiological changes of atrial fibrillation. *Front Physiol* 2020;**11**: 573433.
42. Watanabe K, Arumugam S, Sreedhar R, Thandavarayan RA, Nakamura T, Nakamura M *et al.* Small interfering RNA therapy against carbohydrate sulfotransferase 15 inhibits cardiac remodeling in rats with dilated cardiomyopathy. *Cell Signal* 2015;**27**:1517–24.
43. Blake MR, Parrish DC, Staffenson MA, Sueda S, Woodward WR, Habecker BA. Chondroitin sulfate proteoglycan 4,6 sulfation regulates sympathetic nerve regeneration after myocardial infarction. *eLife* 2022;**11**:e78387.
44. Tsuji-Tamura K, Morino-Koga S, Suzuki S, Ogawa M. The canonical smooth muscle cell marker TAGLN is present in endothelial cells and is involved in angiogenesis. *J Cell Sci* 2021;**134**:jcs254920.
45. Jiao Q, Bai Y, Akaike T, Takeshima H, Ishikawa Y, Minamisawa S. Sarcalumenin is essential for maintaining cardiac function during endurance exercise training. *Am J Physiol Heart Circ Physiol* 2009;**297**:H576–82.
46. Rasi K, Piuhola J, Czabanka M, Sormunen R, Ilves M, Leskinen H *et al.* Collagen XV is necessary for modeling of the extracellular matrix and its deficiency predisposes to cardiomyopathy. *Circ Res* 2010;**107**:1241–52.
47. Mutlak M, Kehat I. Dual specific phosphatases (DUSPs) in cardiac hypertrophy and failure. *Cell Signal* 2021;**84**:110033.
48. Sinner MF, Stepas KA, Moser CB, Krijthe BP, Aspelund T, Sotoodehnia N *et al.* B-type natriuretic peptide and C-reactive protein in the prediction of atrial fibrillation risk: the CHARGE-AF consortium of community-based cohort studies. *Europace* 2014;**16**: 1426–33.
49. Patton KK, Ellinor PT, Heckbert SR, Christenson RH, DeFilippi C, Gottdiener JS *et al.* N-terminal pro-B-type natriuretic peptide is a major predictor of the development of atrial fibrillation. *Circulation* 2009;**120**:1768–74.
50. Chang ICY, Chen LY, Chong JPC, Austin E, Quay CN, Gong L *et al.* Plasma mid-regional pro-atrial natriuretic peptide and N-terminal pro-brain natriuretic peptide improve discrimination of lone atrial fibrillation. *Int J Cardiol* 2015;**188**:10–2.
51. Patton KK, Heckbert SR, Alonso A, Bahrami H, Lima JAC, Burke G *et al.* N-terminal pro-B-type natriuretic peptide as a predictor of incident atrial fibrillation in the Multi-Ethnic Study of Atherosclerosis: the effects of age, sex and ethnicity. *Heart* 2013;**99**:1832–6.
52. Man J, Barnett P, Christoffels VM. Structure and function of the Nppa–Nppb cluster locus during heart development and disease. *Cell Mol Life Sci* 2018;**75**: 1435–44.
53. Chung MK, Martin DO, Sprecher D, Wazni O, Kanderian A, Carnes CA *et al.* C-reactive protein elevation in patients with atrial arrhythmias. *Circulation* 2001;**104**:2886–91.
54. Melas M, Beltsios ET, Adamou A, Koumarelas K, McBride KL. Molecular diagnosis of hypertrophic cardiomyopathy (HCM): in the heart of cardiac disease. *J Clin Med* 2023;**12**:225.
55. Bongini C, Ferrantini C, Girolami F, Coppini R, Arretini A, Targetti M *et al.* Impact of genotype on the occurrence of atrial fibrillation in patients with hypertrophic cardiomyopathy. *Am J Cardiol* 2016;**117**:1151–9.
56. Kaptoge S, Di Angelantonio E, Lowe G, Pepys MB, Thompson SG, Collins R *et al.* C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. *Lancet Lond Engl* 2010;**375**:132–40.
57. Liang F, Wang Y. Coronary heart disease and atrial fibrillation: a vicious cycle. *Am J Physiol Heart Circ Physiol* 2021;**320**:H1–H12.
58. Geelhoed B, Börschel CS, Niiranen T, Palosaari T, Havulinna AS, Fouodo CJK *et al.* Assessment of causality of natriuretic peptides and atrial fibrillation and heart failure: a Mendelian randomization study in the FINRISK cohort. *Europace* 2020;**22**: 1463–9.
59. Tissue expression of NCAM1-summary-the Human Protein Atlas. https://www.proteinatlas.org/ENSG00000149294-NCAM1/tissue#expression_cluster (6 February 2023).
60. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J *et al.* Genomic atlas of the human plasma proteome. *Nature* 2018;**558**:73–9.
61. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR *et al.* Mendelian randomization. *Nat Rev Methods Primer* 2022;**2**:1–21.
62. Phulka JS, Ashraf M, Bajwa BK, Pare G, Laksman Z. Current state and future of polygenic risk scores in cardiometabolic disease: a scoping review. *Circ Genomic Precis Med* 2023; **16**:286–313.