Biobank-Scale Plasma Proteomics Identifies Novel Biomarkers in Hypertrophic Cardiomyopathy

Plasma proteomics for hypertrophic cardiomyopathy

Jonathan H Chan, MBiochem[a,b], Christopher Grace, DPhil[a,b], Mohsen Mazidi, PhD[c] , Robert Clarke, MD[c], Carolyn Y. Ho, MD[d], Stefan Neubauer, MD[a], Christopher M. Kramer, MD[e] on behalf of HCMR Investigators, Hugh Watkins, MD, PhD[a,b], Anuj Goel, MD[a,b]

[a] Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

[b] Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK

[c] Nuffield Department of Population Health, University of Oxford, Oxford, UK

[d] Cardiovascular Division, Department of Medicine and Department of Radiology, Brigham and Women's Hospital, Boston, MA

[e] Cardiovascular Division, University of Virginia Health System, Charlottesville

Corresponding author: Anuj Goel; anuj.goel@cardiov.ox.ac.uk; Centre for Human Genetics, Roosevelt Dr, Headington, Oxford OX3 7BN, UK

Total word count (including Title Page, Abstract, Text, References, Tables and Figures Legends but excluding Abbreviations): 4991

Abstract

Background

Hypertrophic cardiomyopathy (HCM) is characterized by substantial heterogeneity in both clinical phenotype and risk of adverse outcomes including heart failure and sudden cardiac death. This highlights the need for robust biomarkers for risk stratification and while previous studies have identified the role of select plasma proteins, comprehensive large-scale proteomic analyses have been limited in HCM.

Methods

We performed case-control analysis of 2922 plasma proteins in 49,588 UK Biobank participants (100 HCM cases) to identify proteins associated with HCM. External replication analyses were performed in the deCODE Genetics Icelandic study (51 cases/38,904 controls) and All of Us (546 cases/41,049 controls) datasets. Associations with adverse clinical outcomes and cardiac endophenotypes of disease severity were further identified, and causal relationships evaluated using Mendelian randomisation. Relative biomarker importance was also assessed by joint modelling via machine learning.

Results

We identified novel associations between angiopoietin-2 (ANGPT2) and latent transforming growth factor-beta binding protein 2 (LTBP2) with HCM, with both also showing prognostic utility for heart failure-related outcomes in HCM cases. We also confirmed the associations of established

biomarkers (e.g. NTproBNP, troponins I and T) with HCM cases, cardiac imaging markers of disease severity, and adverse outcomes. Mendelian randomisation analyses supported a causal effect of HCM on increasing NTproBNP and troponin T levels.

Conclusions

This biobank-scale plasma proteomic study in HCM identified ANGPT2 and LTBP2 as novel HCM biomarkers with potential diagnostic and prognostic utility. These findings highlight the potential for plasma proteomics to improve risk prediction and provide insight into HCM pathobiology.

Non-standard Abbreviations and Acronyms

ANGPT2: Angiopoietin-2
AoU: All of Us
CI: Confidence interval
CMR: Cardiac magnetic resonance
DCM: Dilated cardiomyopathy
HCMR: Hypertrophic Cardiomyopathy Registry
HF: Heart failure
LTBP2: Latent transforming growth factor-beta binding protein 2
MR: Mendelian randomisation
NPPB: Brain natriuretic peptide
NTproBNP: N-terminal pro-brain natriuretic peptide
TNNI3: Troponin I
TNNT2: Troponin T
UKB: UK Biobank
UKB-PPP: UK Biobank Pharma Proteomics Project

Introduction

Hypertrophic cardiomyopathy (HCM) is one of the most common inherited heart diseases, afflicting 1 in 500 people worldwide[1]. It is characterised by hypertrophy of the left ventricle (LV) which can cause LV dysfunction, heart failure (HF), and sudden cardiac death (SCD)[2]. This potential severity in conjunction with its marked clinical heterogeneity[2] have prompted the need for reliable biomarkers to aid risk stratification. Plasma proteins present an accessible, low-cost modality to evaluate underlying pathological changes and have aided risk prediction for various diseases[3]. In HCM, prior studies have implicated proteins including N-terminal pro-brain natriuretic peptide (NTproBNP) and troponins which reflect pathophysiological processes such as myocardial wall stress and necrosis[4]. Various cohort studies have also demonstrated their associations with imaging-derived endophenotypes of disease severity including myocardial fibrosis[5,6], and with adverse clinical outcomes[7,8].

The advent of biobank-scale plasma proteomic datasets as the UK Biobank Pharma Proteomics Project (UKB-PPP) measuring ~3000 plasma proteins in >50,000 individuals[9] has enabled comprehensive, population-scale analyses. These have identified novel biomarkers and disease mechanisms in cardiac conditions as heart failure, myocardial infarction and coronary artery disease[3,10]. In this study (Figure 1), we leveraged the UKB-PPP to evaluate the association of 2922 proteins with HCM to discover disease-associated plasma protein biomarkers. We further validated a subset of these biomarkers in external replication datasets including the

deCODE Genetics SomaScan V4 study, All of Us (AoU) and HCM Registry (HCMR) cohorts, and demonstrated their prognostic utility through association with adverse clinical outcomes. We also explored their causal relevance for HCM using Mendelian randomisation (MR) and leveraged machine learning (ML)-based classifiers to infer their relative importances in a joint setting. Together, these analyses refine our understanding of the plasma proteomic landscape in HCM, delineating the diagnostic and prognostic utility of both established and novel biomarkers.

Methods

Full methods detailed in the Supplemental Material. Analysis code provided at https://github.com/JonChan0/HCM_Plasma_Proteomics. UKB and AoU data available to approved researchers at https://biobank.ndph.ox.ac.uk/showcase/ and https://www.researchallofus.org/. The National Research Ethics Service approved the UKB (16/NW/0274) and the HCMR studies (14/SC/0190). The National Bioethics Committee of Iceland approved the deCODE study. All patients provided written informed consent.

Results

Discovery association analysis identified known and novel biomarkers of HCM

The UKB-PPP (Table S1) enabled discovery association analysis across 2922 plasma proteins in 49,588 individuals (100 HCM cases). Our case-control differential expression analysis identified 5 proteins significantly associated with prevalent HCM (Figure 2a). These included established biomarkers as NTproBNP, its related brain natriuretic peptide (NPPB), and troponin I (TNNI3)[4]. The remaining 2 biomarkers angiopoietin-2 (ANGPT2) and latent transforming growth factor-beta binding protein 2 (LTBP2) had not previously been associated with HCM disease status. Among the 49,546 controls (including 58 incident HCM cases), we performed time-to-event (TTE) analysis to evaluate associations between baseline protein levels and incident HCM diagnosis in UKB-PPP with median follow-up time 14.9 years [IQR, 14.1 - 15.7 years]. This provided orthogonal support for NTproBNP (HR per SD=2.40 [1.81 – 3.19] 95% CI, $p=1.13 \times 10^{-5}$) (Figure 2b).

Some of the HCM-associated biomarkers had previously been associated with heart failure (HF)[11] so we repeated case-control analyses in the non-HCM HF cohort (Figure S1a). All HCM biomarkers except Troponin I were also associated with HF, indicating that they likely reflect pathobiological changes, including cardiac wall stress and cellular damage, which are known features of all-cause HF. However, the overall signal in HCM was

not driven solely by HF-positive HCM cases because analysis of non-HF HCM cases showed significant associations with 3 of these 5 biomarkers (Figures S2b).

Biomarkers were validated in external deCODE Genetics, All of Us and HCM Registry datasets

The deCODE Genetics SomaScan V4 study provided Icelandic population-scale data for external replication of significant findings from the primary analysis. Of the 5 HCM-associated biomarkers, both known (NTproBNP and Troponin I) and novel (ANGPT2) biomarkers were replicated in this external study with the other 2 (NPPB and LTBP2) not tested due to data unavailability (Figure 3a). This highlights the robustness of these signals especially of the novel biomarker ANPGT2 across 2 distinct population cohorts in the UK and Iceland.

The All of Us cohort (Table S2) also provided population-scale data for known biomarkers NTproBNP, NPPB, troponins I and T. Such proteins replicated albeit with smaller effect sizes (Figure 3b). This may reflect AoU's non-standardised measurements and differences in cohort composition relative to UKB. Sensitivity analysis of the only-HF subgroup replicated the biomarker non-specificity observed in UKB (Figure S1c).

HCMR provided a case-only dataset with joint cardiac magnetic resonance (CMR) imaging and plasma protein data (Table S2). As such, we also investigated cross-phenotype correlations between plasma proteins and CMR imaging-derived LV measures in HCMR. We specifically evaluated

measures of hypertrophy, fibrosis and contractile function. Increased levels of NTproBNP and troponin T associated with greater hypertrophy, fibrosis, and less contractile function (Table 1). These external datasets ultimately supported both novel associations as ANGPT2 and known associations of NTproBNP/troponins with HCM[4], and also provided insight into their link to imaging-derived endophenotypes of severity.

Known (NTproBNP) and novel biomarkers (ANGPT2 & LTBP2) showed prognostic utility in UK Biobank cases

Given these indirect associations with disease severity, we also directly investigated the association of known (NTproBNP) and novel (ANGPT2 & LTBP2) biomarkers with adverse clinical outcomes in UKB cases. Median follow-up times (years) for each of the composite outcomes were 11.2 [IQR, 7.5 – 14.5] for Overall, 14.3 [IQR, 9.7 – 15.4] for HF and 14.7 [13.9 – 15.5] for ventricular arrhythmia (VA) with summary numbers of events detailed in Table S4. Consistent with previous studies[8], NTproBNP was a significant predictor of adverse clinical outcomes including the Overall (HR per SD=2.16 [1.46 – 3.21] 95% CI, $p=1.0 \times 10^{-4}$), HF (HR per SD=3.70 [2.22 – 6.16] 95% CI, $p=4.91 \times 10^{-7}$) and VA (HR per SD=2.52 [1.16 – 5.05] 95% CI, p=0.0093) composites. ANPGT2 and LTBP2 were also associated with incidence of HF composite (HR per SD=1.76 [1.22 – 2.54] 95% CI, p=0.0026) and (HR per SD=1.99 [1.22 – 3.25] 95% CI, p=0.0056) respectively (Figure 4a). Stratification of cases by their plasma protein levels further supported these associations (Figure 4b–d). These TTE analyses validated NTproBNP's prognostic utility and identified ANGPT2

and LTBP2 as novel biomarkers associated with HF-related outcomes in HCM patients.

Mendelian randomisation analyses established causal relationships between HCM and its biomarkers NTproBNP and troponin T

The analyses so far represented observational approaches without clarifying underlying causality. Thus, we also evaluated the causal relevance of these associations using MR. We used common genetic variants associated with phenotypes as instrumental variables to test whether differences in genetically-determined levels of an exposure were associated with an outcome independently of unmeasured confounders[12].

Analyses demonstrated that increased genetic liability to HCM was causally associated with higher plasma levels of NTproBNP and troponin T levels with effects [95% CI] of 0.10 ([0.05-0.16]) and 0.07 ([0.02-0.11]) standard deviations respectively per doubling of liability (Figure 5). Sensitivity analyses (Figure S4, Figure S5) including colocalisation analyses (Figure S6) highlighted the importance of variants at *BAG3* and *MYOZ1/SYNPO2L* loci respectively in driving effects, but these analyses were constrained by the limited power of the HCMR-derived GWASs. From the UKB-derived GWASs of plasma proteins, nominally significant signals were obtained for HCM causing increased levels of NTproBNP and NPPB (Table S5) but these were non-significant after multiple testing correction. The discrepancy for the HCM-to-NTproBNP signal between the HCMR and

UKB-derived GWASs may arise from the UKB's 'healthy volunteer' bias[13] such that for UKB's 'healthier' HCM cases, this may result in lower biomarker levels than the clinically enrolled HCMR cases. As such, the HCM instruments had weaker NTproBNP associations in the UKB GWAS relative to HCMR's (Figure S7). Overall, MR showed that HCM causes elevated NTproBNP and troponin T plasma levels, validating their roles as biomarkers.

Machine learning enabled joint proteomic modelling to evaluate relative predictive importance

Discovery analyses evaluated the relationship between proteins and HCM marginally, with each protein assessed individually. To evaluate these relationships in a joint analysis, we used ML to model proteins jointly, capture potential non-linear effects and ultimately, assess each biomarker's relative importance. SHAP values enabled this by quantifying how much a specific protein's measured level shifts the model's prediction for an individual towards or away from 'case' prediction[14]. As expected, the biomarkers had greater average predictive contribution in cases relative to controls (Figure 6a). Specifically, NTproBNP had the greatest mean |SHAP| value in both groups (cases: 0.85 [0.62–1.07]; controls: 0.041 [0.040–0.042] 95% CI), reflecting its greatest predictive utility in a combined setting. This was mostly driven by high NTproBNP values pushing the model towards 'case' prediction (high positive SHAP value) (Figure 6b/c, Figure S8). Troponin I and ANGPT2 also had significant joint contributions (Figure 6a). Medium-high levels pushed individuals towards

‘case’ prediction albeit with reduced magnitudes relative to NTproBNP. (Figure 6b/c, Figure S8). In a jointly modelled setting, elevated levels of these 3 biomarkers (especially NTproBNP) had substantial additive predictive utility in predicting cases.

Discussion

HCM is characterized by clinical heterogeneity and potential for severe adverse outcomes, motivating a need for biomarkers with predictive and prognostic utility[2]. Leveraging UKB-PPP and external datasets including deCODE Genetics, AoU and HCMR, this study presents a biobank-scale interrogation of the plasma proteome in HCM and highlights both established and novel proteins with potential relevance for diagnosis, prognosis and understanding of latent disease pathways.

These novel biomarkers included ANGPT2 and LTBP2. ANGPT2 is a context-dependent regulator of vascular permeability at the endothelium, controlling the balance between its stable, and its permeable, destabilized state[15]. This is mediated by the TIE1/2 tyrosine kinase receptors. ANGPT2 has been linked to numerous heart conditions. Specifically, elevated plasma ANGPT2 was associated with HF-linked hospitalisation in atrial fibrillation patients[16], as well as with incidence and severity of HF in cardiovascular disease-free and HF cohorts respectively[17]. This likely reflects shared pathobiological mechanisms with HCM such as endothelial stress. Pathological changes in HCM as hypertrophy and fibrosis may cause strain on the microvasculature, resulting in protective activation of

ANPGT2 expression as an agonist of TIE2 receptor to inhibit vascular leakage[15]. However, overexpression of ANGPT2 in mouse hearts resulted in increased endothelial destabilisation and myocardial fibrosis[18] likely via antagonism of TIE2. This context-dependent activity of ANGPT2 confounds our understanding of its role in HCM but regardless, elevated levels associate with disease in population-scale datasets from both the UK and Iceland.

LTBP2 is a secreted glycoprotein which incorporates and stabilises extracellular matrix components such as microfibrils[19]. While it was not present in the deCODE Genetics SomaScan V4 study for external replication, previous studies have identified it as a marker of cardiac fibrosis with upregulated expression and protein localisation in fibrotic regions from HF mouse models and patient-derived tissues[20]. A rat DCM model study further indicated its role in driving fibrosis as siRNA-mediated knockdown reduced fibrosis[19]. Plasma LTBP2 levels have also been correlated with myocardial levels, and event-free survival from SCD in DCM patients[21]. This known biology of LTBP2, and the increased plasma levels we observed in HCM, suggests increased expression of this ECM remodelling protein likely by activated cardiac fibroblasts. As such, plasma LTBP2 may serve as a novel biomarker of HCM reflecting active ECM remodelling and fibrosis.

NTproBNP, troponins I and T are established biomarkers of HCM[4,22]. NTproBNP is produced by proteolysis of pro-brain natriuretic peptide released by cardiomyocytes under wall stress[4]. In contrast, Troponins I and

T plasma levels reflect extracellular release following myocardial injury[4], or even elevated mechanical stress in the absence of injury[23]. They were associated with imaging markers of disease severity[6] and adverse clinical outcomes in HCM patients[7,8,24]. Our analyses also confirmed their disease association at the population-scale and via MR, demonstrated the causal role of HCM in increasing NTproBNP and Troponin T levels. NTproBNP's association with incident disease may also indicate its ability to reflect subclinical, pathological changes prior to clinical presentation as suggested by studies of non-hypertrophic individuals carrying causal sarcomeric mutations[22,25].

We also leveraged ML via SHAP analysis to move beyond marginal associations and assess the joint predictive contributions of biomarkers. This identified the importance of NTproBNP in a multi-marker context, driven by its elevated levels signifying 'case' status. Troponin I and ANGPT2 also had significant additive predictive contributions to 'case' prediction in a joint setting, suggesting the utility of a multi-marker panel whereby each protein contributes distinct information.

A limitation of our analyses was the heterogeneity of datasets such that not all datasets measured the same proteins of interest. For example, limited proteins were profiled in HCMR and AoU datasets, and some were not measured in UKB (Troponin T) and deCODE (LTBP2). However, this deCODE study enabled external validation of most primary findings despite the only modest correlation in protein measurements previously reported between the differing platforms of Olink and SomaScan V4[26]. The

use of targeted panels in UKB and deCODE also likely biases towards those with known disease associations and limits discovery of unknown proteins with potential association. A key limitation was also the low number of HCM cases in UKB which limited the statistical power of our biomarker discovery analyses. We also assumed prevalent case status if individuals were diagnosed up to 5 years after blood collection to capture signal from undiagnosed HCM patients. This does provide a limiting assumption given the lack of ground truth for disease manifestation, but is justified by our sensitivity analysis of this lag time (Figure S9) and previous reports[27].

Conclusions

We leveraged population-scale biobank and clinical cohort data to identify novel (ANGPT2 and LTBP2) and known (e.g. NTproBNP and troponins) plasma protein biomarkers with diverse functional relevance in HCM. We further demonstrated their association with incident diagnoses, cardiac imaging markers of disease severity, and adverse clinical outcomes. This highlights the utility of plasma protein profiling in aiding risk and severity prediction in HCM and paves the way for further studies and mechanistic interrogation to clarify the link between HCM and these novel biomarkers.

Acknowledgements

Authors acknowledge the computational resources provided by the Biomedical Research Computing Facility at the University of Oxford and guidance from Professor Qiang Zhang. SN and HW acknowledge support from the Oxford NIHR Biomedical Research Centre. HCMR was supported by the National Institutes of Health, the National Heart, Lung, and Blood Institute (U01HL117006-01A1), Cytokinetics, and the Thomas Fund. This research has been conducted using the UK Biobank Resource under Application Number 11223. We acknowledge *All of Us* participants for their contributions and thank the NIH's *All of Us* Research Program for making available the examined data. We also acknowledge deCODE Genetics and thank them for the use of their SomaScan V4 study data.

Sources of Funding

J.H.C is supported by the Radcliffe Department of Medicine, Merton College, and Clarendon Fund, University of Oxford, UK. H.W. and A.G. are supported by the British Heart Foundation's Big Beat Challenge award to CureHeart (BBC/F/21/220106).

Disclosures

None.

Supplemental Material

- Supplemental Methods
- Figures S1-S10
- Table S1-S7
- Supplemental .xlsx File

References

1. McKenna WJ, Judge DP. Epidemiology of the inherited cardiomyopathies. *Nat. Rev. Cardiol.* 2021;18:22–36.

2. Olivotto I, Cecchi F, Poggesi C, Yacoub MH. Patterns of disease progression in hypertrophic cardiomyopathy an individualized approach to clinical staging. *Circ. Hear. Fail.* 2012;5:535–546.

3. Schuermans A, Pournamdari AB, Lee J, Bhukar R, Ganesh S, Darosa N, Small AM, Yu Z, Hornsby W, Koyama S, et al. Integrative proteomic analyses across common cardiac diseases yield mechanistic insights and enhanced prediction. *Nat. Cardiovasc. Res.* 2024;3:1516–1530.

4. Matthia EL, Setteducato ML, Elzeneini M, Vernace N, Salerno M, Kramer CM, Keeley EC. Circulating Biomarkers in Hypertrophic Cardiomyopathy. *J. Am. Heart Assoc.* 2022;11:1–15.

5. Ho CY, López B, Coelho-Filho OR, Lakdawala NK, Cirino AL, Jarolim P, Kwong R, González A, Colan SD, Seidman JG, et al. Myocardial Fibrosis as an Early Manifestation of Hypertrophic Cardiomyopathy. *N. Engl. J. Med.* 2010;363:552–563.

6. Neubauer S, Kolm P, Ho CY, Kwong RY, Desai MY, Dolman SF,

Appelbaum E, Desvigne-Nickens P, DiMarco JP, Friedrich MG, et al. Distinct Subgroups in Hypertrophic Cardiomyopathy in the NHLBI HCM Registry. *J. Am. Coll. Cardiol.* 2019;74:2333–2345.

7. Gommans DHF, Cramer GE, Fouraux MA, Heijmans S, Michels M, Timmermans J, Verheugt FWA, de Boer RA, Kofflard MJM, Brouwer MA. Usefulness of High-Sensitivity Cardiac Troponin T to Predict Long-Term Outcome in Patients with Hypertrophic Cardiomyopathy. *Am. J. Cardiol.* 2021;152:120–124.

8. Coats CJ, Gallagher MJ, Foley M, O'Mahony C, Critoph C, Gimeno J, Dawnay A, McKenna WJ, Elliott PM. Relation between serum N-terminal pro-brain natriuretic peptide and prognosis in patients with hypertrophic cardiomyopathy. *Eur. Heart J.* 2013;34:2529–2537.

9. Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, Surendran P, Mahajan A, Robins C, Vasquez-Grinnell SG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature*. 2023;622:329–338.

10. Lind L, Mazidi M, Clarke R, Bennett DA, Zheng R. Measured and genetically predicted protein levels and cardiovascular diseases in UK Biobank and China Kadoorie Biobank. *Nat. Cardiovasc. Res.* 2024;5.

11. 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: Comparing and

Contrasting Their Genetic Determinants and Associations with Outcomes. *Circulation*. 2019;139:2754–2764.

12. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, Hartwig FP, Kutalik Z, Holmes M V., Minelli C, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res.* 2023;4:186.

13. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, Collins R, Allen NE. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. *Am. J. Epidemiol.* 2017;186:1026–1034.

14. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 2017;2017-Decem:4766–4775.

15. Saharinen P, Eklund L, Alitalo K. Therapeutic targeting of the angiopoietin-TIE pathway. *Nat. Rev. Drug Discov.* 2017;16:635–661.

16. Benz AP, Hijazi Z, Lindbäck J, Connolly SJ, Eikelboom JW, Kastner P, Ziegler A, Alexander JH, Granger CB, Lopes RD, et al. Plasma angiopoietin-2 and its association with heart failure in patients with atrial fibrillation. *Europace*. 2023;25:1–10.

17. Peplinski BS, Houston BA, Bluemke DA, Kawut SM, Kolb TM, Kronmal RA, Lima JAC, Ralph DD, Rayner SG, Steinberg ZL, et al. Associations of Angiopoietins With Heart Failure Incidence and Severity. *J. Card. Fail.* 2021;27:786–795.

18. Chen JX, Zeng H, Reese J, Aschner JL, Meyrick B. Overexpression of angiopoietin-2 impairs myocardial angiogenesis and exacerbates cardiac fibrosis in the diabetic db/db mouse model. *Am. J. Physiol. - Hear. Circ. Physiol.* 2012;302:1003–1012.

19. Pang XF, Lin X, Du JJ, Zeng DY. LTBP2 knockdown by siRNA reverses myocardial oxidative stress injury, fibrosis and remodelling during dilated cardiomyopathy. *Acta Physiol.* 2020;228:1–14.

20. Park S, Ranjbarvaziri S, Lay FD, Zhao P, Miller MJ, Dhaliwal JS, Huertas-Vazquez A, Wu X, Qiao R, Soffer JM, et al. Genetic regulation of fibroblast activation and proliferation in cardiac fibrosis. *Circulation*. 2018;138:1224–1235.

21. Nishiura K, Yokokawa T, Misaka T, Ichimura S, Tomita Y, Miura S, Shimizu T, Sato T, Kaneshiro T, Oikawa M, et al. Prognostic Role of Circulating LTBP-2 in Patients With Dilated Cardiomyopathy: A Novel Biomarker Reflecting Extracellular Matrix LTBP-2 Accumulation. *Can. J. Cardiol.* 2023;39:1436–1445.

22. Ho JE, Shi L, Day SM, Colan SD, Russell MW, Towbin JA, Sherrid M V., Canter CE, Jefferies JL, Murphy A, et al. Biomarkers of cardiovascular stress and fibrosis in preclinical hypertrophic cardiomyopathy. *Open Hear.* 2017;4:1–9.

23. Lee MMY, Masri A. Differentiating Cardiac Troponin Levels During Cardiac Myosin Inhibition or Cardiac Myosin Activation Treatments: Drug Effect or the Canary in the Coal Mine? *Curr. Heart Fail. Rep.*

2023;20:504–518.

24. Jansen M, Algül S, Bosman LP, Michels M, van der Velden J, de Boer RA, van Tintelen JP, Asselbergs FW, Baas AF. Blood-based biomarkers for the prediction of hypertrophic cardiomyopathy prognosis: a systematic review and meta-analysis. *ESC Hear. Fail.* 2022;9:3418–3434.

25. Ho CY, Cirino AL, Lakdawala NK, Groarke J, Valente AM, Semsarian C, Colan SD, Orav EJ. Evolution of hypertrophic cardiomyopathy in sarcomere mutation carriers. *Heart*. 2016;102:1805–1812.

26. Eldjarn GH, Ferkingstad E, Lund SH, Helgason H, Magnusson OT, Gunnarsdottir K, Olafsdottir TA, Halldorsson B V., Olason PI, Zink F, et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature*. 2023;622:348–358.

27. Naidu SS, Sutton MB, Gao W, Fine JT, Xie J, Desai NR, Owens AT. Frequency and clinicoeconomic impact of delays to definitive diagnosis of obstructive hypertrophic cardiomyopathy in the United States. *J. Med. Econ.* 2023;26:682–690.

28. Harper AR, Goel A, Grace C, Thomson KL, Petersen SE, Xu X, Waring A, Ormondroyd E, Kramer CM, Ho CY, et al. Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity. *Nat. Genet.* 2021;53:135–142.

29. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, Gunnarsdottir K, Helgason A, Oddsson A,

Halldorsson B V., et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* 2021;53:1712–1721.

30. Bick AG, Metcalf GA, Mayo KR, Lichtenstein L, Rura S, Carroll RJ, Musick A, Linder JE, Jordan IK, Nagar SD, et al. Genomic data in the All of Us Research Program. *Nature*. 2024;627:340–346.

31. Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, Crews DC, Doria A, Estrella MM, Froissart M, et al. New Creatinine- and Cystatin C–Based Equations to Estimate GFR without Race. *N. Engl. J. Med.* 2021;385:1737–1749.

32. Cologne J, Hsu W-L, Abbott RD, Ohishi W, Grant EJ, Fujiwara S, Cullings HM. Proportional hazards regression in epidemiologic follow-up studies: an intuitive consideration of primary time scale. *Epidemiology*. 2012;23:565–573.

33. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203–209.

34. Tadros R, Zheng SL, Grace C, Jordà P, Francis C, West DM, Jurgens SJ, Thomson KL, Harper AR, Ormondroyd E, et al. Large-scale genome-wide association analyses identify novel genetic loci and mechanisms in hypertrophic cardiomyopathy. *Nat. Genet.* 2025;

35. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA,

Ziyatdinov A, Benner C, O'Dushlaine C, Barber M, Boutkov B, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* 2021;53:1097–1103.

36. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408.

Tables

Phenotype Group	**CMR Phenotype**	**Plasma Protein**	**Beta [95% CI]**	**p-value**
Hypertrophy	Left Ventricular Mass Index (SD)	NTproBNP (SD)	0.53 [0.50, 0.57]	**1.24E-150**
	Max LV Wall Thickness (SD)		0.51 [0.47, 0.54]	**1.39E-124**
Contractile Function	Left Ventricular Ejection Fraction (SD)		-0.19 [-0.24, -0.15]	**1.32E-17**

	-Longitudinal Peak Systolic LV Strain- Transmural - Mean (SD)		-0.42 [-0.46, -0.38]	**7.57E-75**
	-Circumferential Peak Systolic LV Strain- Transmural - Mean (SD)		-0.27 [-0.32, -0.23]	**1.90E-30**
	Radial Peak Systolic LV Strain- Transmural - Mean (SD)		-0.14 [-0.19, -0.10]	**1.83E-09**
	Cardiac Index (SD)		0.03 [-0.01, 0.08]	0.19
Fibrosis	Total LGE Over All Segments (SD)		0.27 [0.23, 0.31]	**4.72E-45**
	Extracellular Cytoplasmic Volume Fraction (SD)		0.28 [0.23, 0.33]	**2.03E-30**
Hypertrophy	Left Ventricular Mass Index (SD)	Troponin T (SD)	0.47 [0.43, 0.51]	**3.75E-110**
	Max LV Wall Thickness (SD)		0.35 [0.31, 0.39]	**1.96E-54**
Contractile Function	Left Ventricular Ejection Fraction (SD)		-0.25 [-0.29, -0.21]	**4.21E-28**
	-Longitudinal Peak Systolic LV Strain- Transmural - Mean (SD)		-0.43 [-0.48, -0.39]	**1.21E-78**
	-Circumferential Peak Systolic LV Strain-		s-0.34 [-0.39, -0.30]	**1.93E-47**

	Transmural - Mean (SD)			
	Radial Peak Systolic LV Strain- Transmural - Mean (SD)		-0.17 [-0.21, -0.12]	**2.80E-12**
	Cardiac Index (SD)		0.02 [-0.02, 0.07]	0.36
Fibrosis	Total LGE Over All Segments (SD)		0.32 [0.29, 0.36]	**1.73E-62**
	Extracellular Cytoplasmic Volume Fraction (SD)		0.26 [0.21, 0.31]	**2.48E-25**

Table 1. Cross-phenotype association analyses in HCM Registry of NTproBNP and Troponin T with imaging-derived measures of hypertrophy, fibrosis and contractile function. *LV: left ventricle, SD: standard deviation.*

Figures with Figure Legends

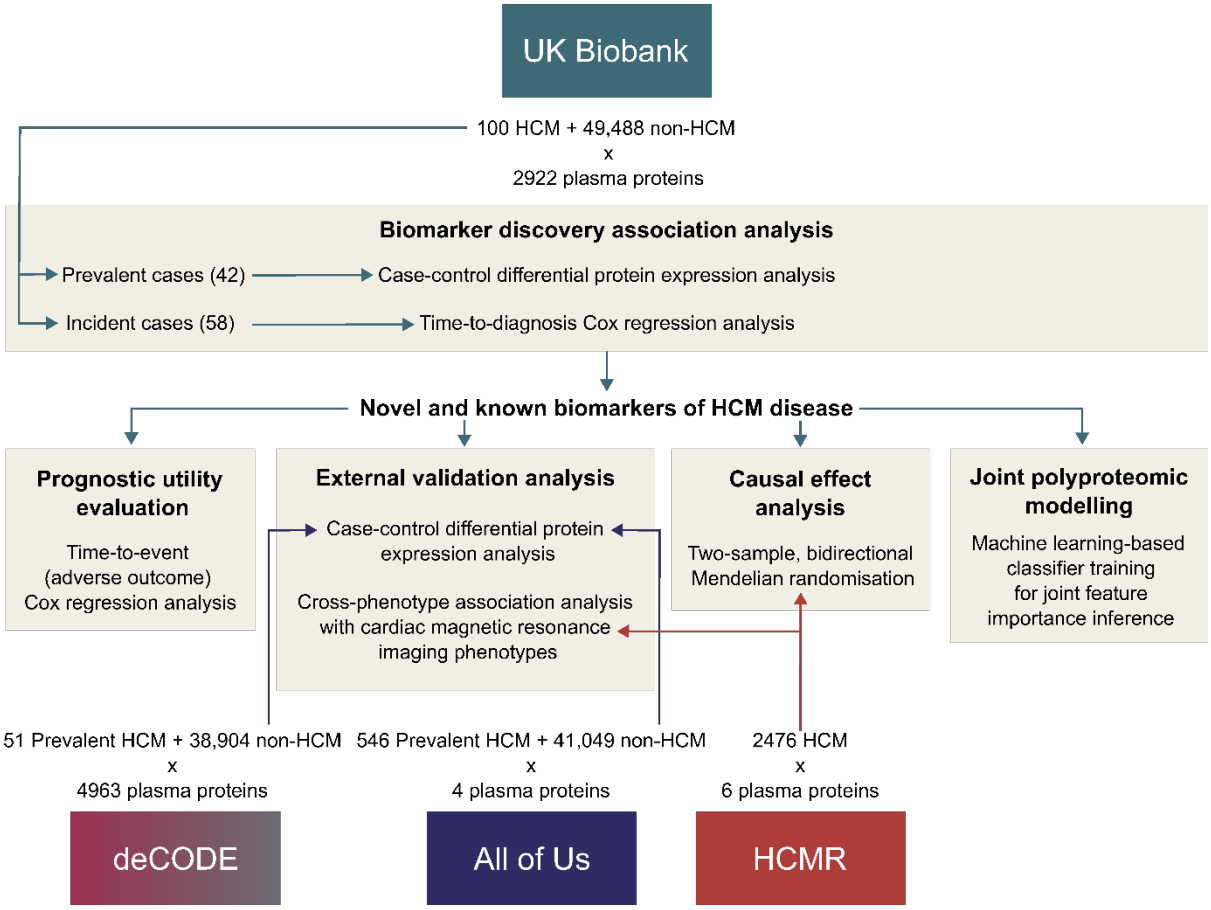

Figure 1. Overview of analyses and datasets involved. *Hypertrophic cardiomyopathy (HCM) cases in UK Biobank, deCODE Genetics, and All of Us datasets were divided into 'prevalent' cases (those diagnosed before the blood sample collection date or within 5 years) and 'incident' cases (those diagnosed after the blood sample collection date + 5 years). This represents an assumed potential lag between disease manifestation and clinical diagnosis. HCMR: Hypertrophic Cardiomyopathy Registry.*

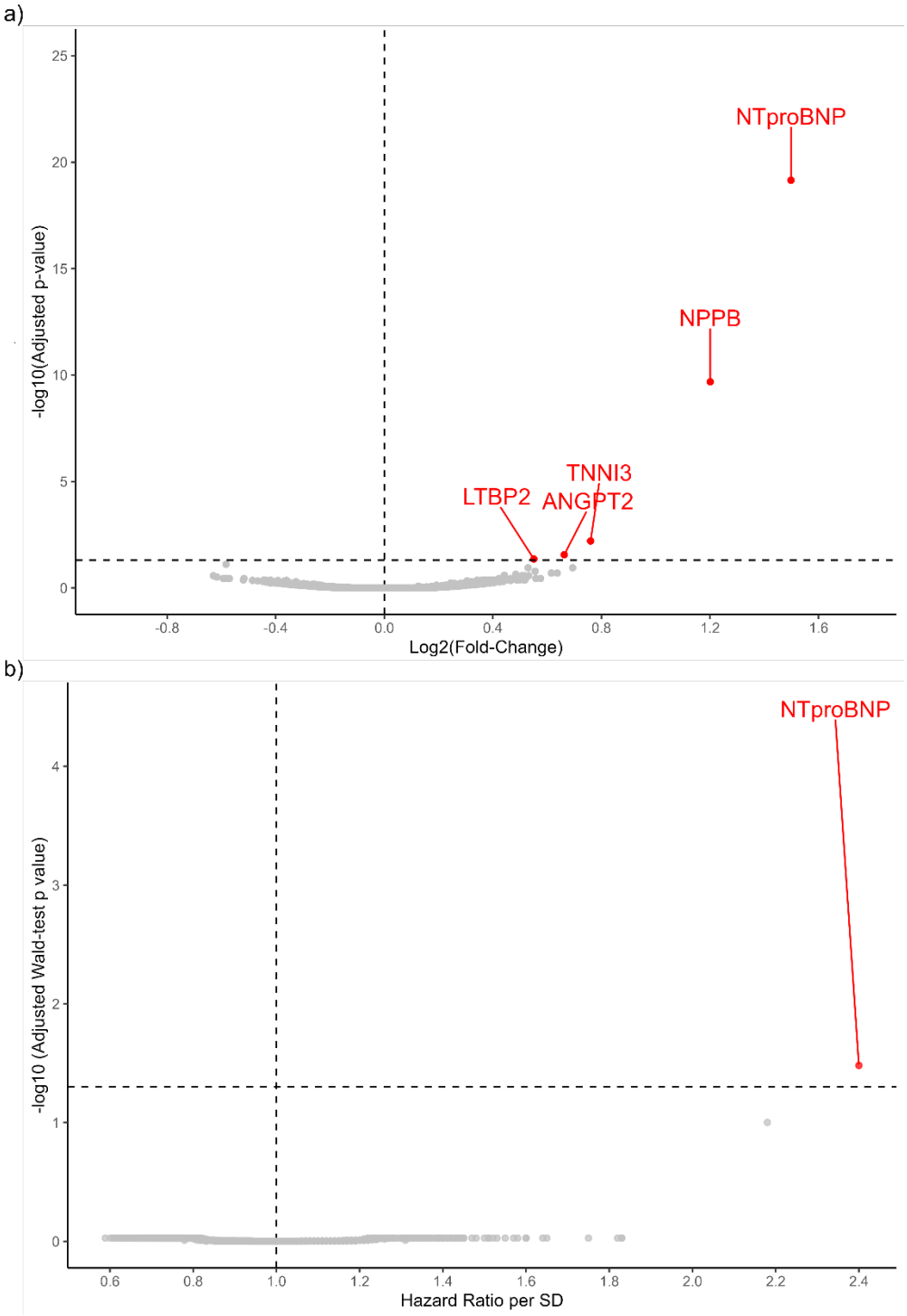

Figure 2. Plasma proteomic association analyses in UK Biobank identified known and novel HCM biomarkers. a) *Case-control differential protein expression analysis using prevalent HCM cases identified previously reported biomarkers of disease (e.g. NTproBNP) and novel ones (ANGPT2 and LTBP2).* ***b)*** *Time-to-event analysis of incident HCM diagnosis in the 'controls' from above analysis (which includes incident HCM cases) provided orthogonal support for NTproBNP's association with disease. Multiple testing correction (MTC) applied via Benjamini-Hochberg procedure to control false discovery rate at 5%. SD: standard deviation.*

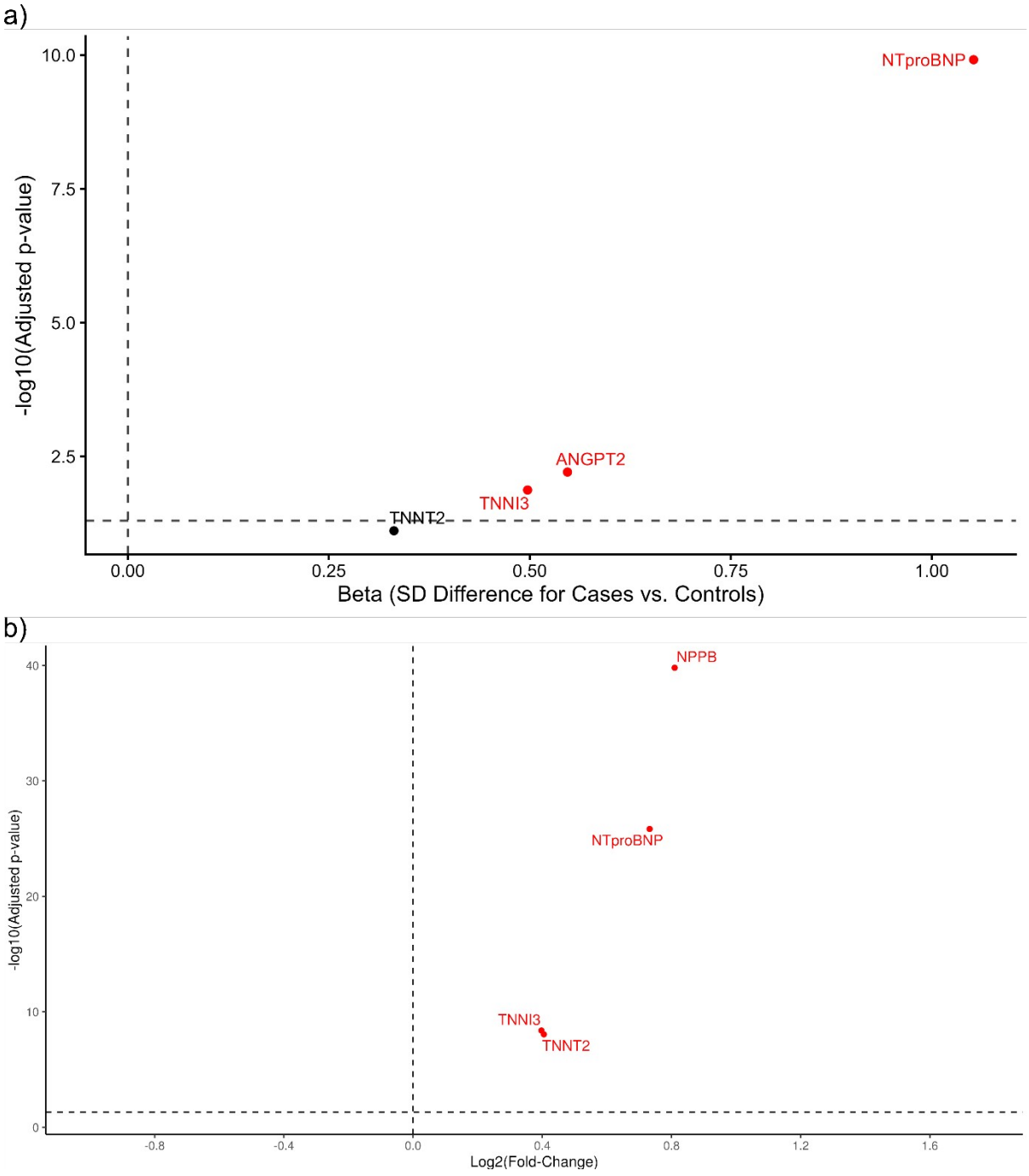

Figure 3. Known and novel biomarkers were validated in external DeCODE Genetics and All of Us datasets. *Case-control differential protein expression analysis in external replication* ***a)*** *DeCODE Genetics and* ***b)*** *All of Us datasets validated known hypertrophic cardiomyopathy (HCM) associations of biomarkers NTproBNP and Troponin I (TNNI3), and novel biomarker ANGPT2 (angiopoietin-2). Plasma proteins tested in external replication DeCODE Genetics dataset included all those significant at the 5% FDR threshold (Figure 2a) if they were also measured in the SomaScan V4 study. Multiple testing correction (MTC) applied via Benjamini-Hochberg procedure to control FDR at 5% (dashed horizontal line).*

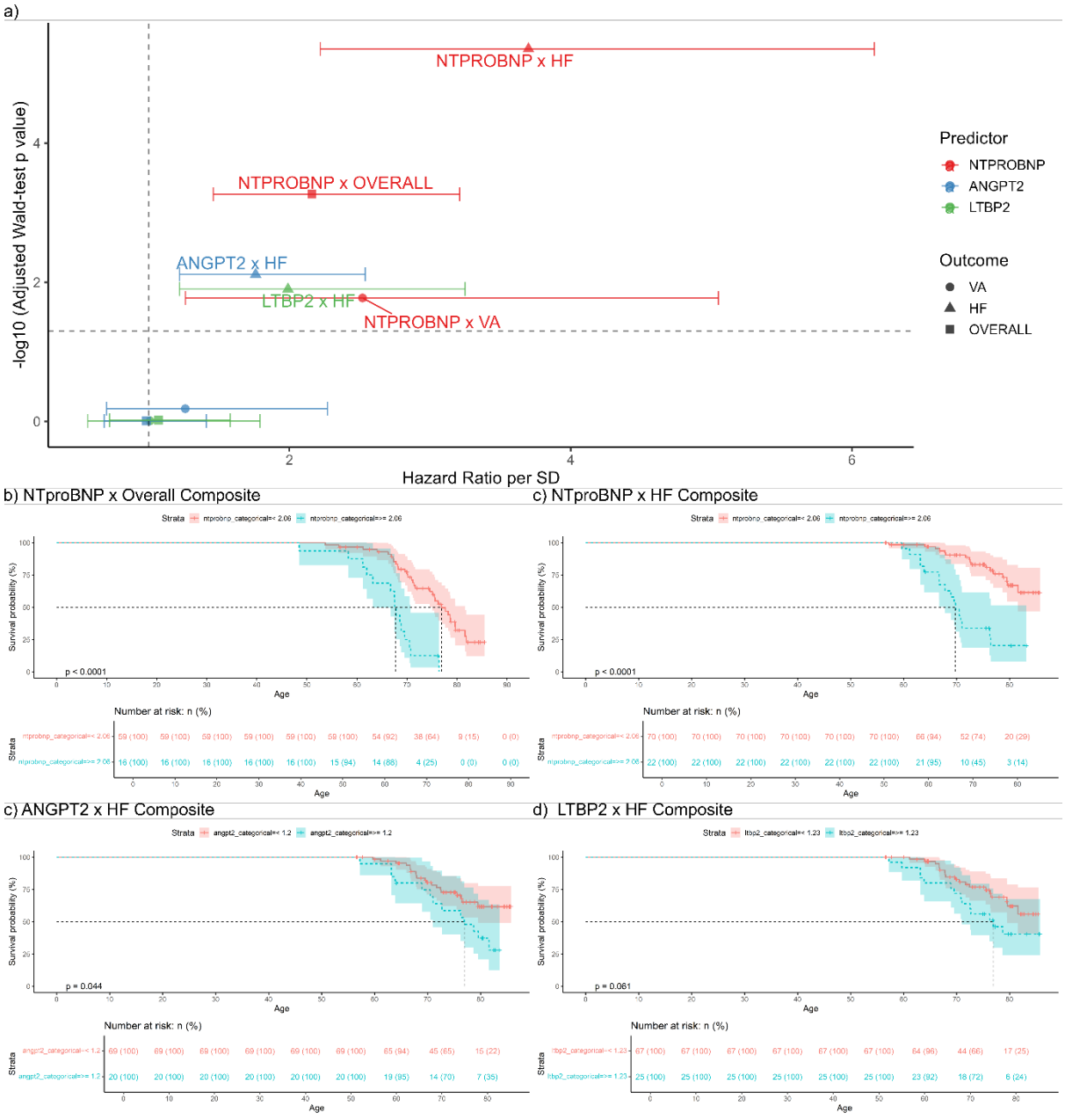

Figure 4. Known and novel biomarkers showed potential prognostic utility. a) *Multivariable Cox regression with incident adverse clinical outcomes in UK Biobank showed associations of known (NTproBNP) and novel (ANGPT2 and LTBP2) biomarkers with composite outcomes after covariate adjustment. Error bars indicate 95% confidence intervals. Multiple testing correction applied via Benjamini-Hochberg procedure to control false discovery rate at 5%.* ***b-d)*** *Kaplan-Meier event curves with stratification by* ***b/c)*** *NTproBNP,* ***c)*** *ANGPT2 and* ***d)*** *LTBP2 levels at baseline (threshold at 75th percentile: 2.06, 1.2 and 1.23 SD respectively) showed significant difference in event incidence between <75th percentile and ≥ 75th percentile subgroups for biomarker-composite combinations prior to covariate adjustment. Statistical test for difference in subgroups via log-rank test. HF: heart failure, SD: standard deviation, VA: ventricular arrhythmia.*

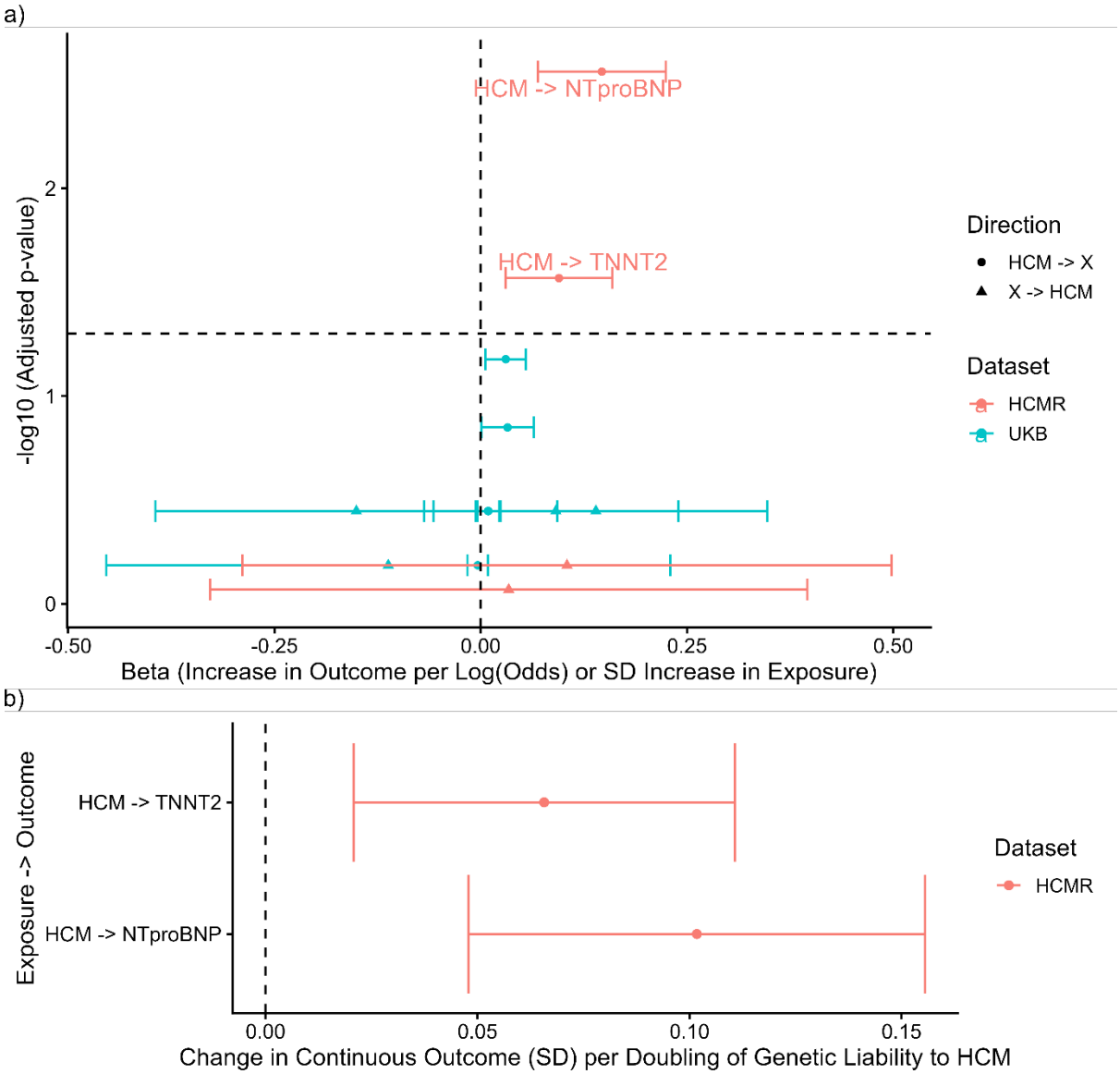

Figure 5. Mendelian randomisation analyses established causal associations between hypertrophic cardiomyopathy (HCM) and known biomarkers. a) *Bidirectional, 2-sample Mendelian randomisation analyses showed significant causality between HCM disease status and plasma proteins (X) prioritised by case-control association analyses. Multiple testing correction (MTC) applied via Benjamini-Hochberg procedure to control FDR at 5%. Associations labelled if significant after MTC.* ***b)*** *Forest plot of significant conclusions indicated that common genetic risk of HCM causes increased NTproBNP and Troponin T (TNNT2) levels. Error bars represent 95% confidence intervals. HCMR: Hypertrophic Cardiomyopathy Registry, UKB: UK Biobank.*

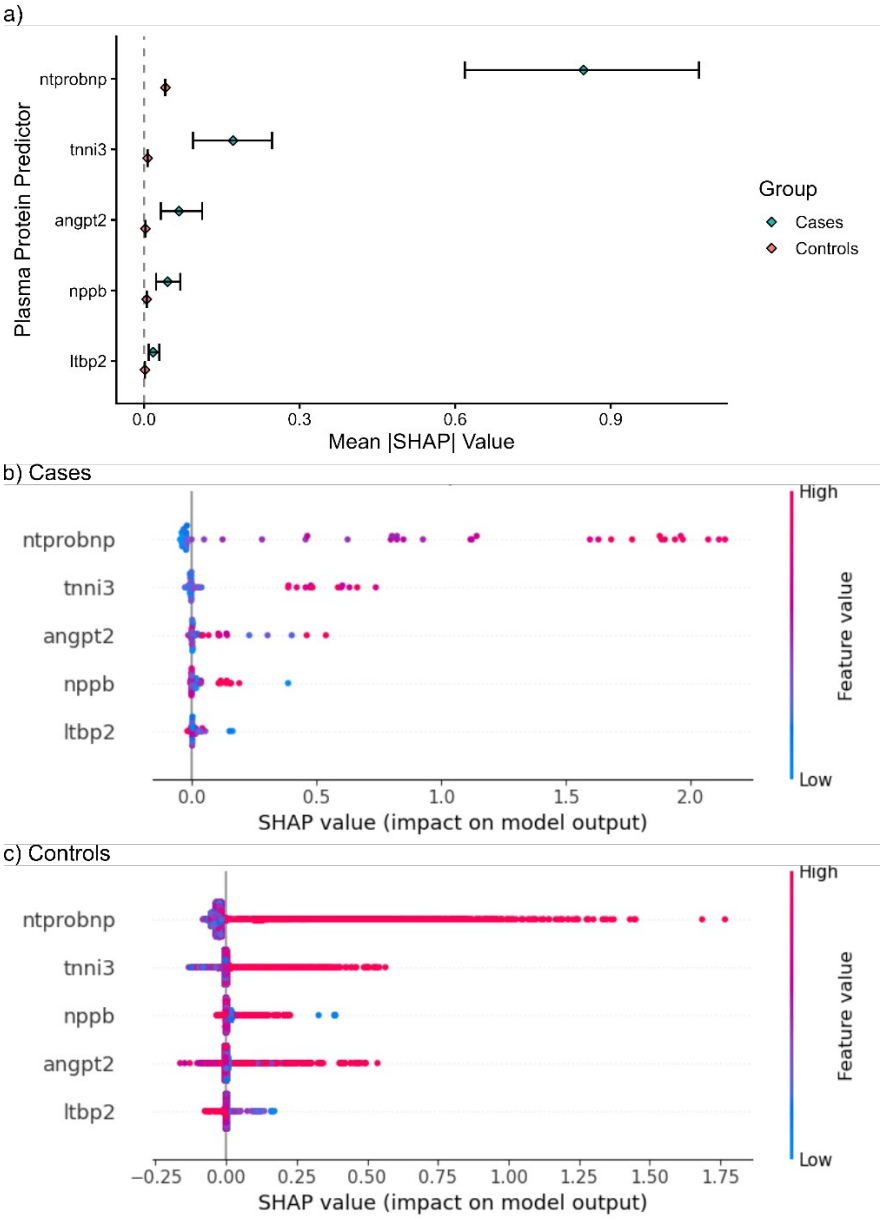

Figure 6. Machine learning-based joint modelling of prioritised plasma proteins highlighted additive predictive contributions of known and novel biomarkers. a) *SHAP values quantify how much a specific protein's measured level pushed the model's prediction for an individual towards or away from a 'case' prediction. Mean absolute SHAP values (unit = log(odds of HCM)) for each plasma protein for prevalent HCM cases and controls showed the relative feature importance and impact on prediction by trained XGBoost classifier. NTproBNP has the greatest predictive contribution in both cases and controls. Error bars represent 95% confidence intervals computed by bootstrapping the evaluated dataset.* ***b/c)*** *Beeswarm plots of SHAP values for each plasma protein stratified for* ***b)*** *cases and* ***c)*** *controls indicates the distribution of SHAP values over all per-individual predictions in the dataset. The extended positive tail for NTproBNP reflects the strong impact of high NTproBNP values for predicting cases whereas its truncated negative tail reflects a weak impact of low NTproBNP values in predicting controls. SHAP: Shapley additive explanation, tnni3: Troponin I, XGBoost: Extreme-Gradient boosted trees.*