









ORIGINAL RESEARCH

Derivation and Validation of Genome-Wide Polygenic Score for Ischemic Heart Failure

Ishan Paranjpe , BS; Noah L. Tsao , BS; Jessica K. De Freitas, BS; Renae Judy, BS; Kumardeep Chaudhary , PhD; Iain S. Forrest , BS; Suraj K. Jaladanki , BS; Manish Paranjpe, BS; Pranav Sharma, BS; CBIPM Genomics Team*; Regeneron Genetics Center; Benjamin S. Glicksberg, PhD; Jagat Narula, MD, PhD; Ron Do , PhD; Scott M. Damrauer , MD; Girish N. Nadkarni , MD, MPH

BACKGROUND: Despite advances in cardiovascular disease and risk factor management, mortality from ischemic heart failure (HF) in patients with coronary artery disease (CAD) remains high. Given the partial role of genetics in HF and lack of reliable risk stratification tools, we developed and validated a polygenic risk score for HF in patients with CAD, which we term HF-PRS.

METHODS AND RESULTS: Using summary statistics from a recent genome-wide association study for HF, we developed candidate PRSs in the Mount Sinai BioMe CAD patient cohort (N=6274) by using the pruning and thresholding method and LDpred. We validated the best score in the Penn Medicine BioBank (N=7250) and performed a subgroup analysis in a high-risk cohort who had undergone coronary catheterization. We observed a significant association between HF-PRS score and ischemic HF even after adjusting for evidence of obstructive CAD in patients of European ancestry in both BioMe (odds ratio [OR], 1.14 per SD; 95% CI, 1.05–1.24; $P=0.003$) and Penn Medicine BioBank (OR, 1.07 per SD; 95% CI, 1.01–1.13; $P=0.016$). In European patients with CAD in Penn Medicine BioBank who had undergone coronary catheterization, individuals in the top 10th percentile of PRS had a 2-fold increased odds of ischemic HF (OR, 2.0; 95% CI, 1.1–3.7; $P=0.02$) compared with the bottom 10th percentile.

CONCLUSIONS: A PRS for HF enables risk stratification in patients with CAD. Future prospective studies aimed at demonstrating clinical utility are warranted for adoption in the patient setting.

Key Words: genomics ■ heart failure ■ personalized medicine ■ polygenic risk score

Despite improvements in cardiovascular prognosis, 5-year mortality from ischemic heart failure (iHF) in coronary artery disease (CAD) is 50%.¹ Because CAD is partially attributable to genetic factors and the genetic heritability of heart failure (HF) is estimated to be 34%,² it is likely that genetic risk is a predictor of development of iHF.

A polygenic risk score (PRS) sums the cumulative weighted contribution of many common genetic variants across the entire genome as a single risk score. Various PRSs have been developed as a marker of genetic predisposition of complex diseases with varied heritability, such as schizophrenia,^{3,4} urinary tract

stones,⁵ and CAD.⁶ Notably, recent work has elucidated the potential role of a CAD PRS for risk stratification beyond contemporary clinical guidelines.⁷ However, there are currently no studies assessing the polygenic contribution to the development of iHF in patients with CAD. Given the paucity of reliable risk prediction tools for incident HF secondary to CAD, a score that incorporates genetic variants provides an orthogonal risk factor to clinical criteria and thus may fill an unmet need.

We aimed to develop a PRS for iHF in a multiethnic biobank from a quaternary care integrated health system and validate it in an external cohort. Because

Correspondence to: Girish Nadkarni, One Gustave Levy Place, Box 1243, New York, NY 10029. E-mail: girish.nadkarni@mountsinai.org

*A complete list of the CBIPM Genomics Team members can be found in the Supplemental Material.

Supplementary Material for this article is available at <https://www.ahajournals.org/doi/suppl/10.1161/JAHA.121.021916>

For Sources of Funding and Disclosures, see pages 6 and 7.

© 2021 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- This is the first validated polygenic risk score that risk stratifies patients with coronary artery disease for risk of heart failure.
- The polygenic risk score associates with heart failure after adjusting for evidence of obstructive coronary artery disease on coronary catheterization.

What Are the Clinical Implications?

- A heart failure polygenic risk score may help risk stratify patients with coronary artery disease and impact clinical decision making after appropriate prospective longitudinal studies.

Nonstandard Abbreviations and Acronyms

iHF	ischemic heart failure
ob-CAD	obstructive coronary artery disease
PC	principal component
PMBB	Penn Medicine BioBank
PRS	polygenic risk score

Individuals with a high polygenic risk burden carry this risk factor throughout life, a PRS may also be used to guide patient management before the onset of clinical disease, thus enabling the early diagnosis, screening, and potentially prevention of HF in a subset of genetically predisposed individuals.

METHODS

The data that support the findings of this study are available from the corresponding author on reasonable request. All clinical and genetic data were collected with approval from the Institutional Review Board with patient consent at each institution.

Identification of Cases

Clinical and demographic data were extracted from an institutional electronic health record database. We identified iHF cases as patients with a diagnosis code for HF with a previous diagnosis of CAD using diagnostic codes for CAD and HF (Table S1). We also identified individuals with evidence of obstructive CAD (ob-CAD) on coronary catheterization. Ob-CAD was defined as the presence of $\geq 50\%$ stenosis in the left main coronary

artery or $\geq 70\%$ stenosis in the left anterior descending, circumflex, or right coronary artery. Individuals who did not undergo coronary catheterization were assumed to not have ob-CAD.

Derivation of PRSs

PRSs measure the cumulative impact of multiple genetic variants on disease risk. For each individual, scores are computed by taking the sum of the dosage of risk variants weighted by their effect on the disease under consideration.

We derived all PRSs using summary statistics from a recent genome-wide association study (GWAS) for all-cause HF.⁸ We computed PRSs using both LDpred⁹ and the pruning and thresholding method. The LDpred algorithm adjusts effect sizes for linkage disequilibrium (LD). We used the 1000 Genomes¹⁰ European population LD reference panel. The tuning parameter in LDpred is ρ , the assumed proportion of causal single-nucleotide polymorphisms. We adjusted ρ between 1 (assumes all single-nucleotide polymorphisms are causal) and 0.001, as done previously.⁶ We then computed a polygenic score for each value of ρ using an additive function, such that $PRS = \sum_i S_i \times G_i$, where S_i = adjusted β statistics for minor allele and G_i = genotype (0, 1, or 2). In the pruning and thresholding method, we varied the P -value threshold from the discovery GWAS study and the r^2 cutoff, a measure of linkage disequilibrium. P values were varied between 0.000001 and 0.1. R^2 values were varied between 0.2 and 0.8. All scores were computed using the PLINK software package.¹¹ We then chose an optimal set of parameters by selecting the score that maximized the association with iHF in Mount Sinai BioMe using logistic regression adjusted for age, sex, and 10 genetic principal components (PCs).

BioMe Discovery Cohort

We used the BioMe Biobank at Mount Sinai as the discovery cohort. BioMe is an electronic health records–linked clinical care cohort composed of >60 000 participants from diverse ancestries (African, Hispanic/Latino, European, and Asian ancestries), with accompanying genome-wide genotyping data for 31 441 participants. Along with the genetic information, BioMe is linked to a wide array of biomedical traits, originating from Mount Sinai's system-wide electronic health records. Enrollment of participants is predominantly through ambulatory care practices and is representative of Mount Sinai's larger patient population. BioMe participants (N=31 441) were genotyped on the Illumina Global Screening Array platform. Quality control and imputation of the Global Screening Array data are detailed in Data S1.

Penn Medicine BioBank Validation Cohort

The Penn Medicine BioBank (PMBB) consists of 60 000 patients recruited from clinical sites across the University of Pennsylvania Health System who have provided consent for access to all electronic health records. This study included a subset of 9973 European ancestry and 5423 African ancestry patients who had undergone genotyping. DNA was extracted from peripheral blood samples within PMBB and genotyped in 3 batches on the Illumina Quad Omni SNP Chip by Regeneron, Global Screening Array V1, and Global Screening Array V2 chips. Following sequencing, standard quality control procedures were applied separately to each batch to remove rare and missing variants as well as variants in linkage disequilibrium. Variants with a minor allele frequency of <0.05 , a missing rate of >0.05 , and a Hardy-Weinberg P of $>10^{-6}$ were removed. Furthermore, samples with a genotype missingness of >0.02 were also removed.

Statistical Analysis

For comparison of demographics and clinical characteristics between cases and controls, we applied a Fisher exact test for categorical variables and a Wilcoxon rank-sum test for continuous variables. PRSs were scaled to have mean 0 and SD 1. For the primary analysis, we associated iHF diagnosis with PRS in individuals with a history of CAD by fitting a logistic regression model, adjusted for age, sex, 10 genetic PCs, history of hypertension and type 2 diabetes, body mass index, history of ob-CAD, and smoking status. Smoking status was defined as being either a current or former smoker based on survey data available in the electronic health record. All analyses were stratified by genetic ancestry groups and included ancestry-specific PCs to account for population stratification. To assess the role of PRS in risk stratification in a high-risk subgroup, we performed a secondary analysis in individuals who had undergone coronary catheterization. In all analyses, statistical significance was determined as $P<0.05$.

RESULTS

Baseline Characteristics

The overall schema of the study is shown in the Figure.

We included patients with a clinical diagnosis of CAD in Mount Sinai BioMe (N=6274) and PMBB (N=7250). Clinical and demographic characteristics are provided in Table 1. In BioMe, we identified 2530 iHF cases and 3744 controls. In BioMe, cases were significantly older (73 versus 71 years; $P=0.002$), had higher mean body mass index (30 versus 29 kg/m²; $P<0.001$), had a higher prevalence of hypertension (83% versus

68%; $P<0.001$) and type 2 diabetes (52% versus 38%; $P<0.001$), and had higher rates of previous cardiac catheterization (31% versus 26%; $P<0.001$). Cases were also more often of African genetic ancestry than controls (35% versus 27%; $P<0.001$).

In the PMBB (Table 1), we observed similar baseline differences in demographics and clinical comorbidities. Cases were significantly older (76 versus 72 years; $P<0.001$), were more commonly male (72% versus 70%; $P=0.018$), and had a higher prevalence of hypertension (92% versus 87%; $P<0.001$) and type 2 diabetes (47% versus 33%; $P<0.001$).

Derivation of PRS for iHF

We created multiple candidate PRSs using summary statistics from a recent GWAS for iHF in primarily

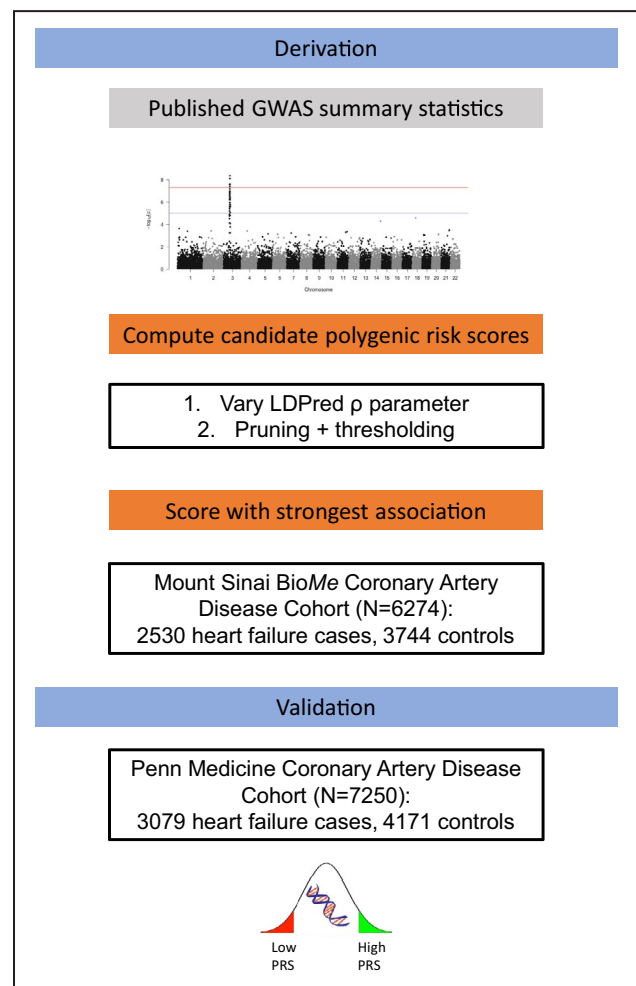


Figure. Derivation and validation of ischemic heart failure polygenic risk score (HF-PRS).

Genome-wide association study (GWAS)⁸ summary statistics were used to derive a candidate PRS using LDpred⁹ and a pruning/thresholding method. The optimal PRS was chosen on the basis of degree of association with heart failure in a training data set, Mount Sinai BioMe, and then external validation was performed in Penn Medicine BioBank.

Table 1. Baseline Demographic and Clinical Characteristics of iHF Cases and Controls in Mount Sinai BioMe and PMBB

Characteristic	Mount Sinai BioMe			PMBB		
	Case (N=2530)	Control (N=3744)	P value	Case (N=3079)	Control (N=4171)	P value
Age, mean (SD), y	72.6 (12)	71.8 (11)	0.002	76.1 (11.9)	71.9 (11.5)	<0.001
BMI, mean (SD), kg/m ²	29.9 (6.8)	29 (5.7)	<0.001	28.7 (6.2)	29.3 (6.3)	<0.001
Ancestry, n(%)						
Hispanic/Latino	651 (25.7)	812 (21.7)	<0.001	0 (0)	0 (0)	
African	890 (35.2)	1011 (27)		637 (20.7)	981 (23.5)	0.004
European	872 (34.5)	1670 (44.6)		2442 (79.3)	3190 (76.5)	0.004
South Asian	92 (3.64)	202 (5.4)		0 (0)	0 (0)	
East Asian	25 (0.988)	49 (1.31)		0 (0)	0 (0)	
Male sex, n (%)	1391 (55)	2040 (54.5)	0.72	2222 (72.2)	2902 (69.6)	0.018
Current or former smoker, n (%)	401 (15.8)	656 (17.5)	0.33	1937 (62.9)	2487 (59.6)	<0.001
Underwent coronary catheterization, n (%)	795 (31.4)	973 (26)	<0.001	678 (22.0)	767 (18.4)	<0.001
Clinical comorbidities, n (%)						
Hypertension	2105 (83.2)	2548 (68)	<0.001	2823 (91.7)	3620 (86.8)	<0.001
Type 2 diabetes	1320 (52.2)	1406 (37.6)	<0.001	1433 (46.5)	1357 (32.5)	<0.001
Family history of heart failure	62 (2.5)	64 (1.7)	<0.001	NA	NA	NA

BMI indicates body mass index; iHF, ischemic heart failure; and PMBB, Penn Medicine BioBank.

European patients and tested their association with iHF in BioMe. Specifically, we applied the pruning and thresholding method by varying the *P* value and *r*² parameters and the LDpred algorithm by varying the *ρ* parameter (Table S2). We selected the derivation strategy that maximized the odds ratio (OR) for association. Applying LDpred with *ρ*=0.001 produced a score (6 214 514 variants) with greatest OR for association with iHF (OR, 1.15; 95% CI, 1.09–1.20) and was used for all downstream analysis.

Association of HF-PRS With iHF

We tested the association of the HF-PRS with ischemic failure (iHF) in patients with CAD by fitting a

logistic regression model, adjusted for age, sex, 10 genetic PCs, history of hypertension and type 2 diabetes, body mass index, history of ob-CAD, and smoking. All analyses were stratified by genetic ancestry groups and included ancestry-specific PCs to account for population stratification. We observed a significant association between HF-PRS score and iHF, even after adjusting for evidence of ob-CAD (Table 2) in patients of European ancestry in both BioMe (OR, 1.14 per SD; 95% CI, 1.05–1.24; *P*=0.003) and PMBB (OR, 1.07 per SD; 95% CI, 1.01–1.13; *P*=0.016). A spline representation of HF risk for varying PRS percentiles is provided in Figure S1. To estimate absolute risk in patients with CAD, we then derived HF probabilities for the 90th and

Table 2. Association of HF-PRS Score and iHF in Mount Sinai BioMe and PMBB

Ancestry group	No. of cases	No. of controls	Odds ratio (95% CI)	P value
Mount Sinai BioMe				
African	890	1011	1.10 (1.0–1.2)	0.05
European	872	1670	1.14 (1.05–1.24)	0.003
Hispanic/Latino	651	812	1.02 (0.91–1.14)	0.8
South Asian	92	202	1.05 (0.79–1.39)	0.8
East Asian	25	49	1.07 (0.52–2.2)	0.9
PMBB				
African	918	637	0.97 (0.88–1.08)	0.6
European	2442	3190	1.07 (1.01–1.13)	0.02

iHF indicates ischemic heart failure; PMBB, Penn Medicine BioBank; and PRS, polygenic risk score.

Table 3. Prevalence and Clinical Impact of a High HF-PRS Score in Unrelated European Individuals With CAD Who Underwent Coronary Catheterization

Group	Top 10% of distribution		Top 5% of distribution		Top 1% of distribution	
	Odds ratio (95% CI)	P value	Odds ratio (95% CI)	P value	Odds ratio (95% CI)	P value
Mount Sinai BioMe European (N=651)	1.3 (0.6–2.7)	0.4	1.4 (0.5–3.4)	0.5	2.3 (0.4–14.0)	0.3
Penn Medicine Biobank European (N=1170)	2.0 (1.1–3.7)	0.02	1.9 (0.9–3.7)	0.09	3.1 (0.8–12.7)	0.1

Odds ratios were calculated by comparing those with a high HF-PRS with those in the bottom 10% of the HF-PRS distribution using a logistic regression model, adjusted for age, sex, 10 genetic principal components, body mass index, and history of hypertension, obstructive CAD, type 2 diabetes, and smoking. CAD indicates coronary artery disease; iHF, ischemic heart failure; and PRS, polygenic risk score.

10th percentile of PRS, while fixing the other covariates at their mean values in the logistic regression model used above. In BioMe, for the 90th percentile, absolute iHF risk was 44.5%; and for the 10th percentile, absolute iHF risk was 35.0%, implying an absolute risk reduction of 9.5%.

We also conducted a subgroup analysis in individuals who had undergone coronary catheterization. Among European individuals with CAD in PMBB who had undergone coronary catheterization, individuals in the top 10th percentile of PRS had a 2-fold increased odds of iHF (OR, 2.0; 95% CI, 1.1–3.7; $P=0.02$) compared with the bottom 10th percentile (Table 3).

We then aimed to assess the impact of a high HF-PRS score by comparing risk of iHF for individuals with a high HF-PRS with those with a low HF-PRS, adjusted for age, sex, 10 genetic PCs, body mass index, history of hypertension, ob-CAD, and type 2 diabetes, and smoking. In BioMe, European individuals with a HF-PRS in the top 10th percentile had a 1.5-fold increased risk of iHF (Table 4) compared with those with an HF-PRS in the bottom 10th percentile. In PMBB, European individuals in the top 10% had a 1.3-fold increased risk of iHF (Table 4) compared with those in the bottom 10th percentile. Association of iHF with HF-PRS percentiles for other ancestral groups are provided in Table S3.

DISCUSSION

Previous work has consistently identified the role of a PRS in risk stratification in CAD.⁶ In this study, we developed a PRS for HF secondary to CAD. More important, because coronary artery occlusion is a significant clinical risk factor for HF,^{12,13} we adjusted all analyses for evidence of ob-CAD, as determined by coronary catheterization. We show several key findings. Compared with the bottom 10th percentile, European patients with CAD with an HF-PRS score in the top 10th percentile have a 1.5-fold increased odds of developing HF in BioMe and a 1.3-fold increased odds in an external validation cohort, PMBB. We observed a stronger effect (OR, 2.0) in high-risk individuals who had undergone coronary catheterization. This effect was stronger in PMBB than in BioMe. One potential reason for this stronger association may be differences in recruitment strategies (recruitment from cardiovascular clinics in PMBB versus from general medicine clinics in BioMe) and indication for coronary catheterization. If patients underwent coronary catheterization for a wider range of indications in PMBB than BioMe, there may be a greater range of ischemic heart disease in the PMBB cohort and thus the PRS would allow greater risk discrimination. However, we do not have information on the indication for coronary catheterization. This difference in effect size also demonstrates

Table 4. Prevalence and Clinical Impact of a High HF-PRS Score in Unrelated European Individuals With CAD

Group (percentile)	No. of cases	No. of controls	Prevalence, %	Odds ratio (95% CI)	P value
Mount Sinai BioMe European					
Top 5% of distribution	65	77	46	1.9 (1.2–3.0)	0.009
Top 10% of distribution	121	163	43	1.5 (1.1–2.3)	0.03
Penn Medicine Biobank European					
Top 5% of distribution	159	204	44	1.3 (0.97–1.8)	0.08
Top 10% of distribution	311	415	43	1.3 (1.0–1.7)	0.048

Odds ratios were calculated by comparing those with a high HF-PRS score with those in the bottom 10% of the HF-PRS distribution using a logistic regression model, adjusted for age, sex, 10 genetic principal components, body mass index, and history of hypertension, obstructive CAD, type 2 diabetes, and smoking. CAD indicates coronary artery disease; iHF, ischemic heart failure; and PRS, polygenic risk score.

the need for cohort-specific model development because baseline risk may differ depending on medical comorbidities and nongenetic contributions. When applying the PRS in a clinical setting, the absolute risk difference is dependent on the overall HF prevalence. We restricted our cohort to only patients with CAD, a relatively high-risk population. In this group of patients, our overall HF prevalence was 40%. However, in another large population-level cohort, UK Biobank, the prevalence of HF within patients with CAD was 17%.⁸ Because the UK Biobank cohort is composed of relatively healthy individuals with fewer comorbidities, the absolute risk difference may be smaller.

We observed the strongest effect sizes in individuals of European ancestry and were only able to replicate the association in individuals of African ancestry in BioMe data and not PMBB data. The GWAS from which our PRS was derived was performed in a European cohort and likely includes several ancestry informative markers. Because putative causative markers may be different across ancestral populations and transferability is limited for polygenic conditions,¹⁴ application of summary statistics from a GWAS performed in European individuals may not be ideal.

Our work provides preliminary evidence of the role of a genome-wide genetic risk score in risk stratification of HF in patients with CAD. Current risk stratification tools only provide prognostic information after the onset of HF rather than determination of HF risk in high-risk patients with CAD. Thus, knowledge of genetic predisposition of HF may help fill an unmet clinical need. Because genetic risk is a nonmodifiable risk factor, a PRS may be measured at any point during a patient's lifetime and thus may be applicable before the onset of CAD. In addition, as more information on modifiable risk factors, such as diet, lifestyle, and other medical comorbidities, is collected for large cohorts, future work may allow for communication of absolute HF risk rather than relative risk. Absolute HF risk is a function of not only genetic factors, such as the PRS presented in our work, but also modifiable factors that may increase or decrease genetic predisposition. For example, recent work from Khera et al demonstrated that adherence to a healthy lifestyle decreased the risk of coronary events by 46% in a group of patients with high polygenic risk for CAD.¹⁵

Our study should be interpreted in light of several limitations. First, given the retrospective nature of the study, coronary catheterization was only performed in a subset of individuals as clinically indicated. Next, similar to previous work in which PRS associates with disease risk more strongly in European than non-European ancestries,^{16–20} our PRS was not significantly associated with iHF diagnosis in non-European ancestral

groups. Thus, because of the paucity of GWAS studies performed in non-European populations, our PRS is applicable only in individuals of European ancestry. However, our results demonstrate the potential utility of a genetic risk score in stratifying HF risk in a high-risk population with CAD, and future ancestry-specific GWAS may make this method broadly applicable.

Future work may aim to apply more sophisticated feature engineering to improve the PRS generation process. The HF-PRS we report is a linear combination of LDPred-adjusted effect sizes. However, using feature selection methods, such as regularization and a nonlinear transformation of input features, may provide greater discriminative ability.

In addition, we aim to apply the HF-PRS as a clinical risk stratification tool and determine its utility in conjunction with appropriate clinical assessment. Because HF is a clinical diagnosis with no established screening guidelines, our tool may be used to identify patients with CAD genetically predisposed to progress to HF and thus require close monitoring, assessment of structural abnormalities, and lifestyle modifications. This remains a testable hypothesis in a prospective randomized study. In addition, the HF-PRS may serve as a nonmodifiable risk factor that can be combined with modifiable risk factors, such as medication compliance, diet, and exercise, in a multivariate model. Large prospective cohort studies with paired genotyping and regular survey data will enable such a study.

ARTICLE INFORMATION

Received April 2, 2021; accepted August 3, 2021.

Affiliations

The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY (I.P., K.C., I.S.F., S.K.J., CBIPM Genomics Team, R.D., G.N.N.); Mount Sinai Clinical Intelligence Center (MSCIC), Icahn School of Medicine at Mount Sinai, New York, NY (I.P., J.K.D.F., K.C., I.S.F., S.K.J., B.S.G., G.N.N.); The Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY (I.P., J.K.D.F., S.K.J., B.S.G., G.N.N.); Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA (N.L.T., R.J., S.M.D.); Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY (J.K.D.F., K.C., I.S.F., B.S.G., R.D.); Division of Health Science and Technology, Harvard Medical School, Boston, MA (M.P.); Drexel University College of Medicine, Philadelphia, PA (P.S.); Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York, NY (J.N.); Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY (G.N.N.); and Renal Program, James J. Peters VA Medical Center at Bronx, New York, NY (G.N.N.).

Sources of Funding

Dr Nadkarni is supported by a career development award from the National Institutes of Health (K23DK107908) and is also supported by R01DK108803, U01HG007278, U01HG009610, and 1U01DK116100. Dr Do is supported by R35GM124836 from the National Institute of General Medical Sciences of the National Institutes of Health and R01HL139865 from the National Heart, Lung, and Blood Institute of the National Institutes of Health and previously an American Heart Association Cardiovascular Genome-Phenome Discovery grant (15CVGSPD27130014).

Disclosures

Dr Nadkarni receives financial compensation as consultant and advisory board member for RenalytixAI, Inc, and owns equity in RenalytixAI. Dr Nadkarni is a scientific cofounder of RenalytixAI. Dr Nadkarni has received operational funding from Goldfinch Bio and consulting fees from BioVie Inc and GLG consulting in the past 3 years. The remaining authors have no disclosures to report.

Supplementary Material

Appendix

Data S1

Table S1–S3

Figure S1

REFERENCES

- Taylor CJ, Ordóñez-Mena JM, Roalfe AK, Lay-Flurrie S, Jones NR, Marshall T, Hobbs FDR. Trends in survival after a diagnosis of heart failure in the United Kingdom 2000–2017: population based cohort study. *BMJ*. 2019;364:l223. doi: 10.1136/bmj.l223
- Lindgren MP, PirouziFard MN, Gustav Smith J, Sundquist J, Sundquist K, Zöller B. A Swedish nationwide adoption study of the heritability of heart failure. *JAMA Cardiol*. 2018;3:703–710. doi: 10.1001/jamacardio.2018.1919
- Vassos E, Di Forti M, Coleman J, Iyegbe C, Prata D, Euesden J, O'Reilly P, Curtis C, Koliakou A, Patel H, et al. An examination of polygenic score risk prediction in individuals with first-episode psychosis. *Biol Psychiatry*. 2017;81:470–477. doi: 10.1016/j.biopsych.2016.06.028
- Shafee R, Nanda P, Padmanabhan JL, Tandon N, Alliey-Rodriguez N, Kalapurakkal S, Weiner DJ, Gur RE, Keefe RSE, Hill SK, et al. Polygenic risk for schizophrenia and measured domains of cognition in individuals with psychosis and controls. *Transl Psychiatry*. 2018;8:78. doi: 10.1038/s41398-018-0124-8
- Paranjpe I, Tsao N, Judy R, Paranjpe M, Chaudhary K, Klarin D, Forrest I, O'Hagan R, Kapoor A, Pfall J, et al. Derivation and validation of genome wide polygenic score for urinary tract stone diagnosis. *Kidney Int*. 2020;98:1323–1330. doi: 10.1016/j.kint.2020.04.055
- Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50:1219–1224. doi: 10.1038/s41588-018-0183-z
- Aragam KG, Dobbyn A, Judy R, Chaffin M, Chaudhary K, Hindy G, Cagan A, Finneran P, Weng L-C, Loos RJF, et al. Limitations of contemporary guidelines for managing patients at high genetic risk of coronary artery disease. *J Am Coll Cardiol*. 2020;75:2769–2780. doi: 10.1016/j.jacc.2020.04.027
- Aragam KG, Chaffin M, Levinson RT, McDermott G, Choi SH, Shoemaker MB, Haas ME, Weng L-C, Lindsay ME, Smith JG, et al. Phenotypic refinement of heart failure in a national biobank facilitates genetic discovery. *Circulation*. 2019;139:489–501. doi: 10.1161/CIRCULATIONAHA.118.035774
- Vilhjálmsdóttir B, Yang J, Finucane H, Gusev A, Lindström S, Ripke S, Genovese G, Loh P-R, Bhatia G, Do R, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet*. 2015;97:576–592. doi: 10.1016/j.ajhg.2015.09.001
- Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74. doi: 10.1038/nature15393
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575. doi: 10.1086/519795
- Gheorghiadu M, Sopko G, De Luca L, Velazquez EJ, Parker JD, Binkley PF, Sadowski Z, Golba KS, Prior DL, Rouleau JL, et al. Navigating the crossroads of coronary artery disease and heart failure. *Circulation*. 2006;114:1202–1213. doi: 10.1161/CIRCULATIONAHA.106.623199
- Gheorghiadu M, Bonow RO. Chronic heart failure in the United States: a manifestation of coronary artery disease. *Circulation*. 1998;97:282–289. doi: 10.1161/01.CIR.97.3.282
- Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. *Cell*. 2019;177:26–31. doi: 10.1016/j.cell.2019.02.048
- Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, Chasman DI, Baber U, Mehran R, Rader DJ, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med*. 2016;375:2349–2358. doi: 10.1056/NEJMoa1605086
- Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, Lee P, Bulik-Sullivan B, Collier DA, Huang H, et al; Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–427. doi: 10.1038/nature13595
- Belsky DW, Moffitt TE, Sugden K, Williams B, Houts R, McCarthy J, Caspi A. Development and evaluation of a genetic risk score for obesity. *Biodemography Soc Biol*. 2013;59:85–100. doi: 10.1080/19485565.2013.774628
- Domingue B, Belsky D, Conley D, Harris KM, Boardman J. Polygenic influence on educational attainment: new evidence from the national longitudinal study of adolescent to adult health. *AERA Open*. 2015;1:2332858415599972. doi: 10.1177/2332858415599972
- Lee JJ, Wedow R, Okbay A, Kong E, Maghziyan O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Karlsson Linnér R, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*. 2018;50:1112–1121. doi: 10.1038/s41588-018-0147-3
- Li Z, Chen J, Yu H, He L, Xu Y, Zhang D, Yi Q, Li C, Li X, Shen J, et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat Genet*. 2017;49:1576–1583. doi: 10.1038/ng.3973

SUPPLEMENTAL MATERIAL

Appendix

The Charles Bronfman Institute of Personalized Medicine (CBIPM) Genomics Team Banner Author List and Contribution Statements

All authors/contributors are listed in alphabetical order.

CBIPM Leadership Team

Noura Abul-Husn, Erwin Bottinger, Judy Cho, Ron Do, Steve Ellis, Omri Gottesman, Yuval Itan, Eimear Kenny, Ruth Loos, Amanda Merkelson, Girish Nadkarni, Aniwaa Owusu-Obeng.

Contribution: All authors contributed to securing funding, study design and oversight.

Sequencing and Lab Operations

Bernadette Liggayu, Amanda Merkelson, Janice Morinigo, Patrick Shanley, Quingbin Song

Contribution: All authors are responsible for DNA extraction, sample handling and tracking, and the library information management system.

Clinical Informatics

Noura Abul-Husn, Lili Chan, Steve Ellis, Omri Gottesman, Arden Moscati, Girish Nadkarni, Rajiv Nadukuru, Aniwaa Owusu-Obeng, Teilman Van Vlick

Contribution: All authors are responsible for analysis needed to produce electronic health record extracted data.

Genome Informatics

Gillian Belbin, Dean Bobo, Kumardeep Chaudhary, Sinead Cullina, Ron Do, Aine Duffy, Amanda Dobbyn, Yuval Itan, Eimear Kenny, Margret Linan, Carla Marquez-luna, Arden Moscati, Ha My, Micheal Preuss, Cigdem Sevim, Stephane Wenric, Ryan Walker, Zhe Wang, Yiming Wu.

Contribution: All authors are responsible for analysis needed to produce exome and genotype data.

Regeneron Genetics Center Banner Author List and Contribution Statements

All authors/contributors are listed in alphabetical order.

RGC Management and Leadership Team

Goncalo Abecasis, Aris Baras, Michael Cantor, Giovanni Coppola, Aris Economides, John D. Overton, Jeffrey G. Reid, Alan Shuldiner.

Contribution: All authors contributed to securing funding, study design and oversight. All authors reviewed the final version of the manuscript.

Sequencing and Lab Operations

Christina Beechert, Caitlin Forsythe, Erin D. Fuller, Zhenhua Gu, Michael Lattari, Alexander Lopez, John D. Overton, Thomas D. Schleicher, Maria Sotiropoulos Padilla, Karina Toledo, Louis Widom, Sarah E. Wolf, Manasi Pradhan, Kia Manoochchri, Ricardo H. Ulloa.

Contribution: C.B., C.F., K.T., A.L., and J.D.O. performed and are responsible for sample genotyping. C.B., C.F., E.D.F., M.L., M.S.P., K.T., L.W., S.E.W., A.L., and J.D.O. performed and are responsible for exome sequencing. T.D.S., Z.G., A.L., and J.D.O. conceived and are responsible for laboratory automation. M.P., K.M., R.U., and J.D.O. are responsible for sample tracking and the library information management system.

Genome Informatics

Xiaodong Bai, Suganthi Balasubramanian, Leland Barnard, Andrew Blumenfeld, Yating Chai, Gisu Eom, Lukas Habegger, Young Hahn, Alicia Hawes, Shareef Khalid, Jeffrey G. Reid, Evan K. Maxwell, John Penn, Jeffrey C. Staples, Ashish Yadav.

Contribution: X.B., A.H., Y.C., J.P., and J.G.R. performed and are responsible for analysis needed to produce exome and genotype data. G.E., Y.H., and J.G.R. provided compute infrastructure development and operational support. S.K., S.B., and J.G.R. provide variant and gene annotations and their functional interpretation of variants. E.M., L.B., J.S., A.B., A.Y., L.H., J.G.R. conceived and are responsible for creating, developing, and deploying analysis platforms and computational methods for analyzing genomic data.

Planning, Strategy, and Operations

Paloma M. Guzzardo, Marcus B. Jones, Lyndon J. Mitnau.

Contribution: All authors contributed to the management and coordination of all research activities, planning and execution. All authors contributed to the review process for the final version of the manuscript.

Data S1.

SUPPLEMENTAL METHODS

Quality control of Mount Sinai BioMe genotyping data

Quality control of the GSA data for $n=635623$ variants was performed stratified by ethnicity category (African American, European American, Hispanic American, Other). We performed sample level quality control based on a number of steps including gender discordance, palindromic SNPs, low coverage, contamination, duplicates, discordance with genotyped data; one individual from every pair of individuals with closer than second degree relatedness was removed. Individuals with an ethnicity-specific heterozygosity rate that surpassed ± 6 standard deviations of the population-specific mean, along with individuals with a call rate of $<95\%$ were removed ($N=684$ participants in total). $N=84$ individuals were then removed for exhibiting persistent discordance between EHR recorded and genetic sex. An additional $N=4$ individuals with phenotypically indeterminate sex were also excluded. A further $N=102$ duplicate individuals were also excluded from downstream analysis. Finally, one of each pair of 28 duplicates were excluded. In total 31,911 passed sample level QC for downstream analysis. Sites with a call rate below 95% were excluded ($n=19253$), along with sites that were seen to significantly violate Hardy-Weinberg equilibrium (HWE) when calculated stratified by ancestry. HWE thresholds varied by ethnicity, specifically we set a threshold of $p < 1e-5$ in African American and European American, or $p < 1e-13$ in Hispanic American ($n=11503$ SNPs in total). Variants with a minor allele frequency $< 5\%$ were removed.

Imputation of genotyping data

All BioMe individuals who were successfully genotyped on GSA chip were subsequently imputed into the 1000 Genomes Phase 3 data release. Genotype data which passed the above quality control filters was phased with SHAPEIT2,1 and imputed to 1000 Genomes Phase 3 reference data using IMPUTE version 2.3.2.2.

Segments of the genome, which were known to harbor gross chromosomal anomalies, were filtered out of the final genotype probabilities files. Imputed sites were excluded if the IMPUTE info score was less than 0.4. The mean IMPUTE info score was 0.873.

Genetic Ancestry

Following imputation, principal component analysis was performed on the imputed genotyping data. In all subsequent statistical analyses, the top 10 genetic principal components (PCs) were included as covariates unless otherwise indicated. Ancestral groups were defined by performing clustering on the top genetic PCs and comparing to 1000 Genomes population groups.

Table S1. ICD codes used for identification of coronary artery disease (CAD) and heart failure (HF) cases.

Condition	ICD Code	Type of Code
Heart Failure	I11.0	ICD10
Heart Failure	I11.9	ICD10
Heart Failure	I13.0	ICD10
Heart Failure	I13.9	ICD10
Heart Failure	I50.*	ICD10
Heart Failure	428	ICD9
Heart Failure	428.1	ICD9
Heart Failure	428.2	ICD9
Heart Failure	428.3	ICD9
Heart Failure	428.4	ICD9
Heart Failure	428.9	ICD9
Coronary Artery Disease	36.11	ICD9
Coronary Artery Disease	36.12	ICD9
Coronary Artery Disease	36.13	ICD9
Coronary Artery Disease	36.14	ICD9
Coronary Artery Disease	36.15	ICD9
Coronary Artery Disease	36.16	ICD9
Coronary Artery Disease	36.17	ICD9
Coronary Artery Disease	36.19	ICD9
Coronary Artery Disease	36.2	ICD9
Coronary Artery Disease	410	ICD9
Coronary Artery Disease	410.0	ICD9
Coronary Artery Disease	410.00	ICD9
Coronary Artery Disease	410.01	ICD9

Coronary Artery Disease	410.02	ICD9
Coronary Artery Disease	410.1	ICD9
Coronary Artery Disease	410.10	ICD9
Coronary Artery Disease	410.11	ICD9
Coronary Artery Disease	410.12	ICD9
Coronary Artery Disease	410.2	ICD9
Coronary Artery Disease	410.20	ICD9
Coronary Artery Disease	410.21	ICD9
Coronary Artery Disease	410.22	ICD9
Coronary Artery Disease	410.3	ICD9
Coronary Artery Disease	410.30	ICD9
Coronary Artery Disease	410.31	ICD9
Coronary Artery Disease	410.32	ICD9
Coronary Artery Disease	410.4	ICD9
Coronary Artery Disease	410.40	ICD9
Coronary Artery Disease	410.41	ICD9
Coronary Artery Disease	410.42	ICD9
Coronary Artery Disease	410.5	ICD9
Coronary Artery Disease	410.50	ICD9
Coronary Artery Disease	410.51	ICD9
Coronary Artery Disease	410.52	ICD9

Coronary Artery Disease	410.6	ICD9
Coronary Artery Disease	410.60	ICD9
Coronary Artery Disease	410.61	ICD9
Coronary Artery Disease	410.62	ICD9
Coronary Artery Disease	410.7	ICD9
Coronary Artery Disease	410.70	ICD9
Coronary Artery Disease	410.71	ICD9
Coronary Artery Disease	410.72	ICD9
Coronary Artery Disease	410.8	ICD9
Coronary Artery Disease	410.80	ICD9
Coronary Artery Disease	410.81	ICD9
Coronary Artery Disease	410.82	ICD9
Coronary Artery Disease	410.9	ICD9
Coronary Artery Disease	410.90	ICD9
Coronary Artery Disease	410.91	ICD9
Coronary Artery Disease	410.92	ICD9
Coronary Artery Disease	411	ICD9
Coronary Artery Disease	411.0	ICD9
Coronary Artery Disease	411.1	ICD9
Coronary Artery Disease	411.8	ICD9
Coronary Artery Disease	411.81	ICD9

Coronary Artery Disease	411.89	ICD9
Coronary Artery Disease	412	ICD9
Coronary Artery Disease	414	ICD9
Coronary Artery Disease	414.0	ICD9
Coronary Artery Disease	414.00	ICD9
Coronary Artery Disease	414.01	ICD9
Coronary Artery Disease	414.02	ICD9
Coronary Artery Disease	414.03	ICD9
Coronary Artery Disease	414.04	ICD9
Coronary Artery Disease	414.05	ICD9
Coronary Artery Disease	414.06	ICD9
Coronary Artery Disease	414.07	ICD9
Coronary Artery Disease	414.2	ICD9
Coronary Artery Disease	414.3	ICD9
Coronary Artery Disease	414.4	ICD9
Coronary Artery Disease	414.8	ICD9
Coronary Artery Disease	414.9	ICD9
Coronary Artery Disease	429.7	ICD9
Coronary Artery Disease	429.79	ICD9
Coronary Artery Disease	I21	ICD10
Coronary Artery Disease	I21.0	ICD10

Coronary Artery Disease	I21.01	ICD10
Coronary Artery Disease	I21.02	ICD10
Coronary Artery Disease	I21.09	ICD10
Coronary Artery Disease	I21.1	ICD10
Coronary Artery Disease	I21.11	ICD10
Coronary Artery Disease	I21.19	ICD10
Coronary Artery Disease	I21.2	ICD10
Coronary Artery Disease	I21.21	ICD10
Coronary Artery Disease	I21.29	ICD10
Coronary Artery Disease	I21.3	ICD10
Coronary Artery Disease	I21.4	ICD10
Coronary Artery Disease	I21.9	ICD10
Coronary Artery Disease	I21.A	ICD10
Coronary Artery Disease	I21.A1	ICD10
Coronary Artery Disease	I21.A9	ICD10
Coronary Artery Disease	I22	ICD10
Coronary Artery Disease	I22.0	ICD10
Coronary Artery Disease	I22.1	ICD10
Coronary Artery Disease	I22.2	ICD10
Coronary Artery Disease	I22.8	ICD10
Coronary Artery Disease	I22.9	ICD10

Coronary Artery Disease	I23	ICD10
Coronary Artery Disease	I23.0	ICD10
Coronary Artery Disease	I23.1	ICD10
Coronary Artery Disease	I23.2	ICD10
Coronary Artery Disease	I23.3	ICD10
Coronary Artery Disease	I23.4	ICD10
Coronary Artery Disease	I23.5	ICD10
Coronary Artery Disease	I23.6	ICD10
Coronary Artery Disease	I23.7	ICD10
Coronary Artery Disease	I23.8	ICD10
Coronary Artery Disease	I24	ICD10
Coronary Artery Disease	I24.0	ICD10
Coronary Artery Disease	I24.1	ICD10
Coronary Artery Disease	I24.8	ICD10
Coronary Artery Disease	I24.9	ICD10
Coronary Artery Disease	I25	ICD10
Coronary Artery Disease	I25.1	ICD10
Coronary Artery Disease	I25.10	ICD10
Coronary Artery Disease	I25.11	ICD10
Coronary Artery Disease	I25.110	ICD10
Coronary Artery Disease	I25.111	ICD10

Coronary Artery Disease	I25.118	ICD10
Coronary Artery Disease	I25.119	ICD10
Coronary Artery Disease	I25.2	ICD10
Coronary Artery Disease	I25.5	ICD10
Coronary Artery Disease	I25.6	ICD10
Coronary Artery Disease	I25.7	ICD10
Coronary Artery Disease	I25.70	ICD10
Coronary Artery Disease	I25.700	ICD10
Coronary Artery Disease	I25.701	ICD10
Coronary Artery Disease	I25.708	ICD10
Coronary Artery Disease	I25.709	ICD10
Coronary Artery Disease	I25.71	ICD10
Coronary Artery Disease	I25.710	ICD10
Coronary Artery Disease	I25.711	ICD10
Coronary Artery Disease	I25.718	ICD10
Coronary Artery Disease	I25.719	ICD10
Coronary Artery Disease	I25.72	ICD10
Coronary Artery Disease	I25.720	ICD10
Coronary Artery Disease	I25.721	ICD10
Coronary Artery Disease	I25.728	ICD10
Coronary Artery Disease	I25.729	ICD10

Coronary Artery Disease	I25.73	ICD10
Coronary Artery Disease	I25.730	ICD10
Coronary Artery Disease	I25.731	ICD10
Coronary Artery Disease	I25.738	ICD10
Coronary Artery Disease	I25.739	ICD10
Coronary Artery Disease	I25.75	ICD10
Coronary Artery Disease	I25.750	ICD10
Coronary Artery Disease	I25.751	ICD10
Coronary Artery Disease	I25.758	ICD10
Coronary Artery Disease	I25.759	ICD10
Coronary Artery Disease	I25.76	ICD10
Coronary Artery Disease	I25.760	ICD10
Coronary Artery Disease	I25.761	ICD10
Coronary Artery Disease	I25.768	ICD10
Coronary Artery Disease	I25.769	ICD10
Coronary Artery Disease	I25.79	ICD10
Coronary Artery Disease	I25.790	ICD10
Coronary Artery Disease	I25.791	ICD10
Coronary Artery Disease	I25.798	ICD10
Coronary Artery Disease	I25.799	ICD10
Coronary Artery Disease	I25.8	ICD10

Coronary Artery Disease	I25.81	ICD10
Coronary Artery Disease	I25.810	ICD10
Coronary Artery Disease	I25.811	ICD10
Coronary Artery Disease	I25.812	ICD10
Coronary Artery Disease	I25.82	ICD10
Coronary Artery Disease	I25.83	ICD10
Coronary Artery Disease	I25.84	ICD10
Coronary Artery Disease	I25.89	ICD10
Coronary Artery Disease	I25.9	ICD10
Coronary Artery Disease	33510	CPT
Coronary Artery Disease	33511	CPT
Coronary Artery Disease	33512	CPT
Coronary Artery Disease	33513	CPT
Coronary Artery Disease	33533	CPT
Coronary Artery Disease	33534	CPT
Coronary Artery Disease	33535	CPT
Coronary Artery Disease	33536	CPT
Coronary Artery Disease	92920	CPT
Coronary Artery Disease	92921	CPT
Coronary Artery Disease	92924	CPT
Coronary Artery Disease	92925	CPT

Coronary Artery Disease	92928	CPT
Coronary Artery Disease	92929	CPT
Coronary Artery Disease	92933	CPT
Coronary Artery Disease	92934	CPT
Coronary Artery Disease	92937	CPT
Coronary Artery Disease	92938	CPT
Coronary Artery Disease	92941	CPT
Coronary Artery Disease	92943	CPT
Coronary Artery Disease	92944	CPT
Coronary Artery Disease	92973	CPT
Coronary Artery Disease	92975	CPT

Table S2. Association of candidate polygenic risk scores with ischemic heart failure in Mount Sinai BioMe.

Derivation Strategy	Tuning Parameter(s)	OR per SD (95% Confidence Interval)
Pruning and Thresholding	$p < 0.5$ and $r^2 = 0.2$	1.047 (1.00-1.10)
Pruning and Thresholding	$p < 0.00001$ and $r^2 = 0.2$	1.028 (0.98-1.07)
Pruning and Thresholding	$p < 0.05$ and $r^2 = 0.2$	1.047 (1.00-1.10)
Pruning and Thresholding	$p < 0.5$ and $r^2 = 0.2$	1.047 (1.00-1.10)
Pruning and Thresholding	$p < 0.00001$ and $r^2 = 0.2$	1.028 (0.98-1.07)
Pruning and Thresholding	$p < 0.05$ and $r^2 = 0.2$	1.047 (1.00-1.10)
Pruning and Thresholding	$p < 0.1$ and $r^2 = 0.4$	1.055 (1.01-1.11)
Pruning and Thresholding	$p < 0.000001$ and $r^2 = 0.4$	1.039 (0.99-1.09)
Pruning and Thresholding	$p < 0.001$ and $r^2 = 0.4$	1.055 (1.01-1.11)
Pruning and Thresholding	$p < 0.1$ and $r^2 = 0.4$	1.055 (1.01-1.11)
Pruning and Thresholding	$p < 0.000001$ and $r^2 = 0.4$	1.039 (0.99-1.09)
Pruning and Thresholding	$p < 0.001$ and $r^2 = 0.4$	1.055 (1.01-1.11)
Pruning and Thresholding	$p < 0.05$ and $r^2 = 0.6$	1.085 (1.03-1.14)
Pruning and Thresholding	$p < 0.5$ and $r^2 = 0.6$	1.085 (1.03-1.14)
Pruning and Thresholding	$p < 0.00001$ and $r^2 = 0.6$	1.054 (1.01-1.10)
Pruning and Thresholding	$p < 0.05$ and $r^2 = 0.6$	1.085 (1.03-1.14)
Pruning and Thresholding	$p < 0.5$ and $r^2 = 0.6$	1.085 (1.03-1.14)
Pruning and Thresholding	$p < 0.00001$ and $r^2 = 0.6$	1.054 (1.01-1.10)
Pruning and Thresholding	$p < 0.001$ and $r^2 = 0.8$	1.083 (1.03-1.14)
Pruning and Thresholding	$p < 0.1$ and $r^2 = 0.8$	1.083 (1.03-1.14)

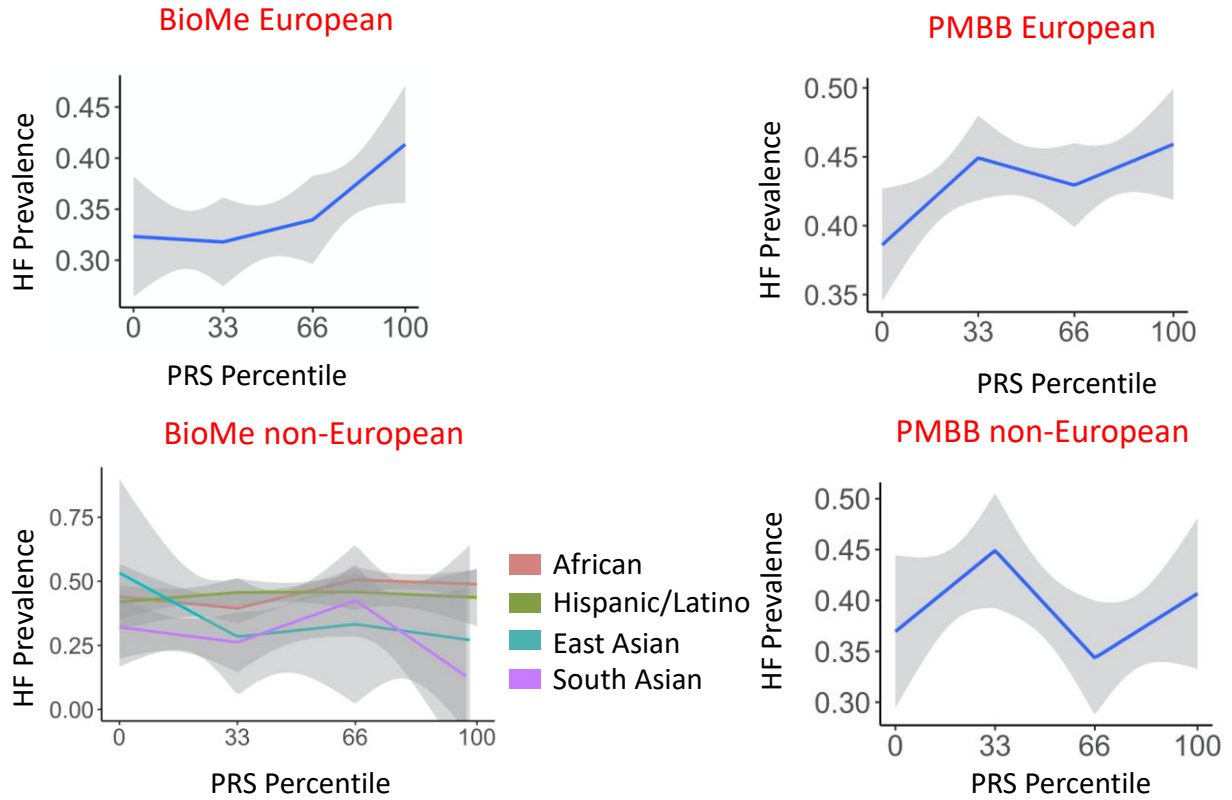
Pruning and Thresholding	$p < 0.000001$ and $r^2 = 0.8$	1.041 (0.99-1.09)
Pruning and Thresholding	$p < 0.001$ and $r^2 = 0.8$	1.083 (1.03-1.14)
Pruning and Thresholding	$p < 0.1$ and $r^2 = 0.8$	1.083 (1.03-1.14)
Pruning and Thresholding	$p < 0.000001$ and $r^2 = 0.8$	1.041 (0.99-1.09)
LDPred Algorithm	$\rho = 1$	1.133 (1.08-1.19)
LDPred Algorithm	$\rho = 0.1$	1.133 (1.08-1.19)
LDPred Algorithm	$\rho = 0.01$	1.137 (1.09-1.19)
LDPred Algorithm	$\rho = 0.001$	1.152 (1.10-1.21)
LDPred Algorithm	$\rho = 0.3$	1.133 (1.08-1.19)
LDPred Algorithm	$\rho = 0.03$	1.134 (1.08-1.19)
LDPred Algorithm	$\rho = 0.003$	1.146 (1.09-1.20)

Table S3. Clinical impact of a high HF-PRS individuals with CAD.

Ancestral Group	Group (percentile)	Number of cases	Number of Controls	Odds Ratio (95% Confidence Interval)	P value
Mount Sinai BioMe					
Hispanic	Top 1% of distribution	8	7	1.9 (0.6 - 6.3)	2.7E-01
	Top 5% of distribution	32	48	0.98 (0.5- 1.8)	9.5E-01
	Top 10% of distribution	63	84	0.95 (0.59 - 1.5)	8.4E-01
African	Top 1% of distribution	6	14	0.54 (0.2- 1.5)	2.7E-01
	Top 5% of distribution	48	48	1.4 (0.8 - 2.3)	2.3E-01
	Top 10% of distribution	96	95	1.4 (0.9 - 2.1)	4.9E-02
South Asian	Top 1% of distribution	0	3	NA	NA
	Top 5% of distribution	3	12	0.5 (0.08 - 2.6)	4.4E-01
	Top 10% of distribution	7	23	0.6 (0.2 - 1.9)	3.6E-01
Penn Medicine Biobank					
African	Top 1% of distribution	9	8	1.8 (0.6 - 5.6)	0.32
	Top 5% of distribution	35	46	0.8 (0.4 - 1.5)	0.45
	Top 10% of distribution	66	96	0.8 (0.5 - 1.3)	0.34

Odds ratios were calculated by comparing those with a high HF-PRS with those in the bottom 10% of the HF-PRS distribution using a logistic regression model adjusted for age, sex, 10 genetic PCs, BMI, and history of HTN, ob-CAD, T2D, and smoking. East Asian ancestry individuals were not analyzed in Mount Sinai BioMe due to low cohort size.

Figure S1. Spline representation of PRS and HF risk.



Spline regression was performed for scaled PRS values in each ancestral group separately. Cutpoints at 33rd and 67th percentile were chosen.