Identification and correction for collider bias in a genome-wide association study of diabetes-related heart failure

Authors

Yan V. Sun, Chang Liu, Qin Hui, ..., the Million Veteran Program, Jacob Joseph, Lawrence S. Phillips

Correspondence

yan.v.sun@emory.edu

This study demonstrated the bidirectional relationship between type 2 diabetes and heart failure, addressed the impact of collider bias on the genome-wide association study of diabetes-related heart failure, and identified and replicated two genetic loci. These discoveries offer pivotal insights into the intricate mechanisms and comorbidities associated with heart failure.

Sun et al., 2024, The American Journal of Human Genetics *111*, 1–13
July 11, 2024
https://doi.org/10.1016/j.ajhg.2024.05.018

Please cite this article in press as: Sun et al., Identification and correction for collider bias in a genome-wide association study of diabetes-related heart failure, The American Journal of Human Genetics (2024), https://doi.org/10.1016/j.ajhg.2024.05.018

ARTICLE

Identification and correction for collider bias in a genome-wide association study of diabetes-related heart failure

Yan V. Sun,[1,2,*] Chang Liu,[2] Qin Hui,[1,2] Jin J. Zhou,[3,4] J. Michael Gaziano,[5,6] Peter W.F. Wilson,[1,7] the Million Veteran Program, Jacob Joseph,[8,9] and Lawrence S. Phillips[1,7]

Summary

Type 2 diabetes (T2D) is a major risk factor for heart failure (HF) and has elevated incidence among individuals with HF. Since genetics and HF can independently influence T2D, collider bias may occur when T2D (i.e., collider) is controlled for by design or analysis. Thus, we conducted a genome-wide association study (GWAS) of diabetes-related HF with correction for collider bias. We first performed a GWAS of HF to identify genetic instrumental variables (GIVs) for HF and to enable bidirectional Mendelian randomization (MR) analysis between T2D and HF. We identified 61 genomic loci, significantly associated with all-cause HF in 114,275 individuals with HF and over 1.5 million controls of European ancestry. Using a two-sample bidirectional MR approach with 59 and 82 GIVs for HF and T2D, respectively, we estimated that T2D increased HF risk (odds ratio [OR] 1.07, 95% confidence interval [CI] 1.04–1.10), while HF also increased T2D risk (OR 1.60, 95% CI 1.36–1.88). Then we performed a GWAS of diabetes-related HF corrected for collider bias due to the study design of index cases. After removing the spurious association of *TCF7L2* locus due to collider bias, we identified two genome-wide significant loci close to *PITX2* (chromosome 4) and *CDKN2B–AS1* (chromosome 9) associated with diabetes-related HF in the Million Veteran Program and replicated the associations in the UK Biobank. Our MR findings provide strong evidence that HF increases T2D risk. As a result, collider bias leads to spurious genetic associations of diabetes-related HF, which can be effectively corrected to identify true positive loci.

Introduction

Heart failure (HF) is a complex, life-threatening syndrome that results from structural and functional impairment of ventricular filling or output. HF affects more than 64 million people worldwide,[1] including 6 million adults in the US.[2] HF prevalence in the US is projected to increase 46% from 2012 to 2030, resulting in over 8 million adults (≥18 years old) with HF.[3] In addition to high mortality and morbidity, HF is also associated with high health care costs with an estimated annual expenditure of $70 billion in the US by 2030.[2]

Type 2 diabetes (T2D) is a complex disease affecting multiple organ systems. The prevalence of T2D has been growing for the past two decades, with age-adjusted prevalence of 9.5% in 1999–2002 and 12% in 2013–2016 among US adults. About 537 million adults live with diabetes around the world, most in low- and middle-income countries (IDF Diabetes Atlas: https://diabetesatlas.org/). HF is one of the most severe diabetes complications affecting T2D individuals' clinical outcomes and quality of life.[4] Observational studies have consistently demonstrated an increased risk of HF in individuals with diabetes compared with those without diabetes, across demographic groups. Even among individuals without T2D, higher levels of fasting glucose and hemoglobin A1c (HbA1c) were associated with increased risk of HF hospitalization.[5,6] The complex pathogenesis of HF in T2D can include toxic effect of hyperglycemia, diabetic cardiomyopathy, coronary microvascular dysfunction, and other comorbid conditions.[7] T2D is associated with a high incidence of both HF with reduced ejection fraction (HFrEF) and HF with preserved ejection fraction (HFpEF),[8] regardless of heterogeneous etiologies, clinical manifestation, and outcomes between HF subtypes. Clinical trials have shown that anti-diabetic medications such as sodium-glucose cotransporter-2 (SGLT2) inhibitors can reduce the risk for HF and subtypes.[9–11] Therefore, further understanding of the genetic and molecular mechanisms of diabetes-related HF may lead to therapeutic targets of HF and HF subtypes.

Genome-wide association studies (GWASs) are designed to identify genetic loci associated with disease and traits by surveying genome-wide single-nucleotide polymorphisms (SNPs). Although no genetic variants have been associated with diabetes-related HF, recent GWASs have identified dozens of loci associated with all-cause HF[12–14] and clinical subtypes, including HFrEF and HFpEF.[12] Two European ancestry-based GWASs of all-cause HF identified 11 and 20 genome-wide significant (GWS) loci using 47,309 and 51,571 individuals with HF, respectively.[12,14]

[1]Atlanta VA Healthcare System, Decatur, GA, USA; [2]Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA, USA; [3]Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA; [4]Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, USA; [5]Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA, USA; [6]Division of Aging, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; [7]Emory University School of Medicine, Atlanta, GA, USA; [8]VA Providence Healthcare System, Providence, RI, USA; [9]The Warren Alpert Medical School of Brown University, Providence, RI, USA
*Correspondence: yan.v.sun@emory.edu
https://doi.org/10.1016/j.ajhg.2024.05.018.

Please cite this article in press as: Sun et al., Identification and correction for collider bias in a genome-wide association study of diabetes-related heart failure, The American Journal of Human Genetics (2024), https://doi.org/10.1016/j.ajhg.2024.05.018

A multi-ancestry GWAS of all-cause HF including 115,150 individuals with HF and over 1.5 million controls identified a total of 47 GWS loci.[13] Genetic and familial studies also estimated the heritability of HF (h^2 ranging from 22% to 34%)[12,15] and several diabetes complications, including diabetes-related cardiovascular disease (h^2 ~18%) and diabetes-related stroke (h^2 ~14%),[16] which suggested substantial genetic contribution to diabetes-related HF yet to be discovered.

When both exposure (e.g., genetic variants) and outcome (e.g., HF) independently influence a common third variable (e.g., T2D), collider bias can occur when the third variable (i.e., collider) is controlled for by design or analysis. Thus, the bidirectional relationship between T2D and HF can introduce collider bias in the GWAS of diabetes-related HF (both genetics and HF affect T2D). Mendelian randomization (MR) methods can support such bidirectional relationship using proper instrumental variables of T2D and HF. MR uses genetic variants robustly associated with exposures or risk factors of interest as genetic instrumental variables (GIVs) to estimate the causal and de-confounded relationship between the exposure or risk factor with the disease outcome.[17] Recent genetic studies identified significant but weak MR association between T2D and all-cause HF using two-sample MR approach (odds ratio [OR] 1.05–1.08),[12,14] compared to observational studies of T2D and HF. Additionally, the estimated genetic correlation between T2D and all-cause HF of 47.3%[14] cannot be fully explained by the moderate MR association between T2D (exposure) and HF (outcome). On the other hand, the hypothesis that HF increases the risk of T2D hasn't been examined in the MR framework, limited by strong GIVs for HF from large independent samples. In the present study, we identified a large set of GIVs of HF from a GWAS meta-analysis, including 114,275 individuals with HF and 1,506,896 controls of European ancestry. Then we performed a large scale GWAS of diabetes-related HF and corrected for collider bias using summary statistics of T2D GWAS to eliminate spurious associations. We also examined and corrected for potential collider bias in diabetes-adjusted GWAS of HF.

Subjects and methods

Study samples

The design of the Million Veteran Program (MVP) has been previously described.[18] Veterans were recruited from over 60 Veterans Health Administration (VHA) (Veterans Affairs [VA]) healthcare systems nationwide since 2011. The MVP has detailed phenotyping through linking the large biobank to an extensive electronic health record (EHR) database from 2003 onward that integrates multiple elements such as diagnosis codes, procedure codes, laboratory values, and imaging reports. All MVP participants were genotyped as part of the study design. MVP has received ethical and study protocol approval by the VA Central Institutional Review Board in accordance with the principles outlined in the Declaration of Helsinki. Informed consent is obtained from all participants to provide blood for genomic analysis and access to their full EHR within the VA prior to and after enrollment. The UK Biobank (UKB) is a prospective study with over 500,000 participants aged 40–69 years recruited in 2006–2010 with extensive phenotypic and genotypic data.[19] The UKB was approved by the North West Multi-centre Research Ethics Committee.

Phenotypic data

In the MVP, individuals with HF were identified as those with an International Classification of Diseases (ICD)-9 code of 428.x or ICD-10 code of I50.x and an echocardiogram performed within 6 months of diagnosis (median time period from diagnosis to echocardiography was 3 days, interquartile range 0–32 days).[12] Based on our previous work, the requirement for echocardiogram improved the specificity of HF diagnosis. The index diagnosis of HF was documented during an outpatient encounter in the majority of participants with HF. We utilized a natural language processing tool to extract left ventricular ejection fraction (LVEF) from the VA Text Integration Utilities documents including values measured within and outside the VA.[12,20,21] Non-HF controls excluded MVP participants with any recorded HF codes at any time based on their EHR data. Diabetes was defined by both (1) either $\geq$1 use of the ICD-9 code 250.xx at a primary care provider visit or $\geq$2 uses of the code in any setting and (2) an outpatient prescription of a diabetes drug based on use of VHA national drug codes.[22]

In the UKB, we defined HF as the presence of self-reported HF, pulmonary edema, or cardiomyopathy at any visit or an ICD-10 or ICD-9 billing code indicative of heart/ventricular failure or a cardiomyopathy of any cause, as described and validated previously, and consistent with that used in recent GWASs of all-cause HF.[12,23] Similar to the MVP definition, non-HF controls excluded participants with any self-reported HF or recorded HF codes at any time. T2D was defined by the primary and secondary ICD-9 (250 diabetes mellitus, juvenile type excluded) and ICD-10 diagnosis codes (E11 non-insulin-dependent diabetes mellitus, E12 malnutrition-related diabetes mellitus, E13 other specified diabetes mellitus, E14 unspecified diabetes mellitus[24]), and self-reported T2D at enrollment.

Genomic data

DNA extracted from participants' blood was genotyped using a customized Affymetrix Axiom biobank array. The array was enriched for both common and rare genetic variants of clinical significance in different ethnic backgrounds. Genotype calling, quality-control procedures, and genotype imputation were previously described.[25] We excluded duplicate samples, samples with more heterozygosity than expected, an excess (>2.5%) of missing genotype calls, or discordance between genetically inferred sex and phenotypic gender.[25] In addition, one individual from each pair of related individuals (more than second-degree relatedness as measured by the KING software[26]) were removed. Prior to imputation, variants that were poorly called (genotype missingness >5%), that weren't in Hardy-Weinberg equilibrium (p value $< 10^{-20}$), or that deviated from their expected allele frequency (>20%) observed in the 1000 Genomes reference data were excluded. After pre-phasing using EAGLE v2.4,[27] we then imputed to the 1000 Genomes phase 3 version 5 reference panel (1000G) using Minimac4.[28] Imputed variants with poor imputation quality ($r^2 < 0.3$) were excluded from further analyses.

The MVP participants were assigned to mutually exclusive racial/ethnic groups using harmonized ancestry and race/ethnicity (HARE), a machine learning algorithm that integrates genetically

inferred ancestry (GIA) with self-identified race/ethnicity (SIRE) as previously described.[29] Briefly, HARE uses GIA to refine SIRE for genetic association studies in three ways: identifies individuals whose SIRE are likely inaccurate, reconciles conflicts among multiple SIRE sources, and imputes missing racial/ethnic information when the predictive confidence is high. HARE assigned >98% of participants with genotype data to one of four non-overlapping groups: non-Hispanic European, non-Hispanic African, Hispanic, and non-Hispanic Asian Americans. The present GWAS of diabetes-related HF focused on the MVP European ancestry.

To replicate the significant loci associated with diabetes-related HF, we performed a similar genetic association analysis in the UKB participants of European ancestry with available genomic data. Additional sample exclusions were implemented for third-degree or closer relatedness (UKB Data Field 22020 includes unrelated participants for the calculation of principal components), sex chromosome aneuploidy, and excess missingness or heterozygosity, as defined by the UKB.

All-cause HF GWAS meta-analysis

Imputed and directly measured genetic variants from the MVP European participants were tested for association assuming an additive genetic model using PLINK2. The GWAS scan included variants with minor allele frequency higher than 1%. Logistic regression of all-cause HF was adjusted for age, sex, and the top ten genotype-derived principal components. We meta-analyzed summary statistics of previously published HF GWAS from the MVP (43,344 individuals with HF, 258,943 controls),[12] HERMES (47,309 individuals with HF, 930,014 controls),[14] and FinnGen (23,622 individuals with HF, 317,939 controls)[30] studies, which included non-overlapping participants of European ancestry using the random-effect meta-analysis model implemented in GWAMA (genome-wide association meta analysis).[31] GWAS results were summarized using FUMA, a platform that annotates, prioritizes, visualizes and interprets GWAS results.[32] GWS SNPs ($p < 5 \times 10^{-8}$) were grouped into a genomic locus based on either $r^2 > 0.1$ or distance between loci of <500 kb using the 1000 Genomes European reference panel. Lead SNPs were defined within each locus if they were independent ($r^2 < 0.1$). We considered loci as novel if the sentinel SNP was of genome-wide significance ($p < 5 \times 10^{-8}$) and located >1 Mb from previously reported GWS SNPs associated with HF.

Based on the meta-analysis summaries of HF, we employed multivariate gene-based analysis of genome-wide association studies (MAGMA)[33] to conduct gene and gene-set analysis by aggregating genetic signals within individual genes, thus revealing gene-based associations that extend beyond the single-marker level. We also conducted tissue expression analysis on 54 distinct tissue types using the Genotype-Tissue Expression[34] (GTEx) dataset, which offers extensive data on gene expression across a diverse array of human tissues, encompassing various organs and biological systems. Additionally, we conducted an analysis of 30 broader tissue categories, omitting specific subtypes or regions, in order to gain insights into gene expression patterns within major tissues. Further, we delved into the functional significance of genes linked to HF through gene set analysis and tissue enrichment analysis, employing the data-driven expression-prioritized integration for complex traits (DEPICT)[35] tool. We applied the false discovery rate[36] (FDR), and associations with a corrected q value <0.2 were deemed statistically significant.

We additionally conducted a transcriptome-wide association study to explore the relationship between gene expression and HF loci using the software FUSION,[37] based on the reference datasets obtained from GTEx[34] V8, including gene expression profiles across tissues including coronary artery, tibial artery, atrial appendage, left ventricle, and skeletal muscle. We then performed colocalization analysis using the coloc[38,39] package in R to identify shared regions between gene expression and HF. Five hypotheses were evaluated for the colocalization analysis: (1) there is no association between the gene expression and HF; (2) there is significant association between the gene expression and HF, but this association is driven solely by the functional effects of the gene expression; (3) there is significant association between the gene expression and HF, but this association is driven solely by the genetic variants identified in HF GWAS; (4) both the gene expression and HF have independent associations with different genetic variants; and (5) there is evidence for colocalization, indicating that the gene expression and HF signals share common causal variant. Colocalization was defined as maximum posterior probability of a sharing causal variant between the gene expression and HF association >0.75.

Additionally, for the identification of overlapping enhancer regions potentially associated with HF, we employed the EnhancerAtlas 2.0 database,[40] which encompasses 295 enhancers specific to various human tissues and cells. For the GWS HF loci, we obtained tissue-specific cis expression quantitative trait loci (eQTL) analysis results from the GTEx[34] version 7 database (https://gtexportal.org) based on the expression data of the following 5 tissue types: coronary artery, tibial artery, atrial appendage, left ventricle, and skeletal muscle. Genes with at least one SNP in cis significantly associated at FDR of ≤0.05 were included. For each gene, a specific threshold for nominal p values was computed. Variants with a nominal p value below the gene-specific threshold were identified significant *cis*-eQTL.[41]

GIVs for T2D and all-cause HF

We selected independent genetic loci ($r^2 < 0.1$) associated with T2D from the large GWAS among participants of European ancestry only or multiple ancestries with predominantly European ancestry participants by 2017.[42] A total of 85 independent T2D-associated SNPs were selected, including SNPs that are GWS ($p < 5 \times 10^{-8}$) in at least one published GWAS of European ancestry but not necessarily GWS in the 2017 T2D GWAS.[42] Among the 85 SNPs, 82 were also present in the all-cause HF GWAS meta-analysis and thus were used as the GIVs of T2D in the downstream MR analysis. From the all-cause HF GWAS meta-analysis described above, we identified independent GWS SNPs as the GIVs for HF in the bi-directional MR analysis.

Two-sample bidirectional MR

Two-sample MR was conducted to examine possible bidirectional causal associations between T2D and all-cause HF using GIVs from previous GWAS of T2D[42] and a large meta-analysis of all-cause HF in the present study. To minimize sample overlap in the two-sample MR design, we used summary statistics of T2D GWAS without UKB and MVP samples and all-cause HF GWAS from the MVP, HERMES, and FinnGen studies, all from studies of European ancestry. We estimated the MR association between T2D and all-cause HF using three complementary methods: inverse-variance weighted (IVW), median weighted, and MR-Egger regression, as implemented in the R package *TwoSampleMR*. We reported IVW estimates when the evidence of pleiotropy was not present. MR-Egger regression was used to identify the horizontal pleiotropy indicated by significant intercept of the

Please cite this article in press as: Sun et al., Identification and correction for collider bias in a genome-wide association study of diabetes-related heart failure, The American Journal of Human Genetics (2024), https://doi.org/10.1016/j.ajhg.2024.05.018

Table 1. Characteristics of the European American participants in the MVP included in the GWAS of diabetes-related and diabetes-adjusted GWAS of heart failure

	All (n = 434,089)		Diabetes (n = 106,321)		Non-diabetes (n = 327,768)	
	HF (n = 68,059)	Non-HF controls (n = 366,030)	HF (n = 31,346)	Non-HF controls (n = 74,975)	HF (n = 36,713)	Non-HF controls (n = 291,055)
Age, years (SD)	69.59 (9.628)	62.49 (14.11)	68.76 (8.44)	66.06 (9.80)	70.29 (10.48)	61.57 (14.88)
Male, n (%)	65884 (96.80)	336064 (91.81)	30509 (97.33)	71454 (95.30)	35375 (96.36)	264610 (90.91)
BMI, kg/m^2 (SD)	31.14 (6.70)	29.30 (5.54)	33.33 (6.80)	31.80 (5.99)	29.26 (6.01)	28.66 (5.23)
Obesity, (BMI ≥30)	35106 (51.58)	145631 (39.79)	20778 (66.29)	43555 (58.09)	14328 (39.03)	102076 (35.07)
Atrial fibrillation, n (%)	22681 (33.33)	23703 (6.48)	10232 (32.64)	6378 (8.51)	12449 (33.91)	17325 (5.95)
Coronary artery disease, n (%)	45739 (67.20)	79154 (21.63)	23091 (73.66)	26373 (35.18)	22648 (61.69)	52781 (18.13)
Chronic kidney disease, n (%)	23610 (34.69)	34404 (9.40)	13615 (43.43)	13472 (17.97)	9995 (27.22)	20932 (7.19)
Diabetes n (%)	31346 (46.06)	74975 (20.48)	31346 (100)	74975 (100)	0 (0)	0 (0)
Hyperlipidemia, n (%)	59567 (87.52)	244010 (66.66)	29724 (94.83)	67780 (90.40)	29843 (81.29)	176230 (60.55)
Hypertension, n (%)	62596 (91.97)	228065 (62.31)	30573 (97.53)	67346 (89.82)	32023 (87.23)	160719 (55.22)

HF, heart failure; SD, standard deviation; n, number; BMI, body mass index.

regression (p value < 0.05). A random-effects model was used to estimate the MR association between exposure and outcome variables for IVW and MR-Egger regression. MR-PRESSO (Mendelian randomization pleiotropy residual sum and outlier) was used to detect and remove outlier GIVs to correct for potential horizontal pleiotropy.[43] As we only evaluate the relationship between T2D and HF, we considered nominal p value of 0.05 as suggestive evidence for MR association.

Latent heritable confounder MR (LHC-MR) is a method designed for analyzing GWAS summary statistics to estimate bidirectional causal effects while accounting for potential heritable confounder between a pair of traits.[44] LHC-MR can overcome the limitations of traditional MR, including under-exploitation of genome-wide markers, sensitivity to the presence of a heritable confounder, and potential sample overlap.[44] LHC-MR extends the traditional MR model by using a structural equation model incorporating the presence of a latent heritable confounder and estimates its contribution to T2D and all-cause HF separately, while simultaneously estimating the bidirectional causal effect between T2D and all-cause HF. We applied this method to estimate the bidirectional relationship between T2D and all-cause HF using summary statistics from a large T2D GWAS[42] and the meta-analysis of HF GWAS, both in European ancestry.

GWAS of diabetes-related HF and diabetes-adjusted HF

We conducted a GWAS of diabetes-related HF using all-cause HF individuals and controls[12] among 106,321 diabetes individuals of European ancestry from the MVP cohort (Table 1). Among them, a total of 31,346 are HF individuals with comorbid T2D, and 74,975 are non-HF diabetes controls. The genetic association of diabetes-related HF was adjusted for age, sex, and top 10 principal components (PCs). Using the same statistical model, we also performed the GWAS of diabetes-related HF among 26,431 unrelated diabetes individuals of European ancestry from the UKB, including 3,506 individuals developed HF. To explore the potentially similar collider bias in diabetes-adjusted HF, we also conducted GWASs of all-cause HF adjusted for T2D status among 434,089 MVP participants of European ancestry, adjusted for age, sex, T2D status, and top 10 PCs.

Correction for collider bias using Slope-Hunter for GWAS of diabetes-related HF

The Slope-Hunter method was developed for correcting collider bias in conditional GWAS using genetic effects of the collider (i.e., T2D) and the outcome variable (i.e., HF).[45] The method employs model-based clustering to identify and utilize variants that specifically affect T2D to estimate an adjustment factor under the assumption that these variants explain more variability in T2D compared to other variant clusters. The method was implemented in the *Slope-Hunter* R package (https://github.com/Osmahmoud/SlopeHunter). We obtained GWAS summary statistics for diabetes-related HF and T2D from the MVP study and considered 7,700,660 variants (minor allele frequency [MAF] >0.01) present in both datasets. An independent set of SNPs was obtained after performing lingkage disequilibrium (LD) pruning using PLINK2 software (r^2 threshold of 0.1 within 250 SNP windows) using the European ancestry population of the 1000 Genomes reference panel. The threshold of $p < 0.001$ was used to define SNP-T2D associations and to fit the main model-based clustering.

Correction for collider bias using instrument effect regression method for GWAS of diabetes-related HF

Under the assumption that the direct genetic effects on HF are independent of those on T2D, we additionally used the instrument effect regression[46] (IER) method to correct for the index event bias of the diabetes-related HF GWAS. The analysis was performed using R package *indexevent*, using the aforementioned independent set of SNPs after LD pruning and the improved version of the simulation extrapolation (SIMEX) algorithm[47,48] to estimate the bias term with 10 simulations performed in each stage of the SIMEX adjustment.

Correction for collider bias for GWAS of diabetes-adjusted HF

Similarly, we performed sensitivity analysis for the GWS loci of the diabetes-adjusted HF GWAS identified using FUMA.[32] Both Slope-Hunter and IER methods were applied to correct for the bias. In

Please cite this article in press as: Sun et al., Identification and correction for collider bias in a genome-wide association study of diabetes-related heart failure, The American Journal of Human Genetics (2024), https://doi.org/10.1016/j.ajhg.2024.05.018

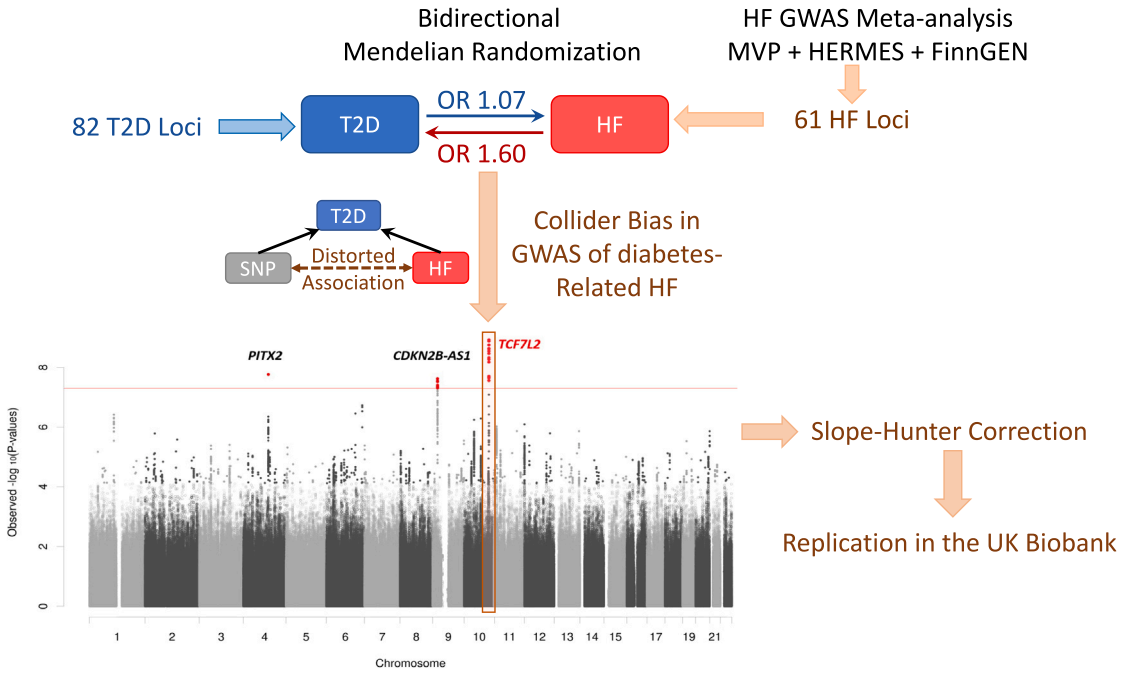

Figure 1. Overview of study design
In the GWAS of diabetes-related HF, *PITX2* and *CDKN2B–AS1* are the two GWS loci; *TCF7L2* highlighted in red is the GWS locus due to collider bias.

addition, we used mtCOJO[49] to estimate the genetic effects on HF conditioning on T2D.

Results

The present study consists of a large meta-analysis of all-cause HF in the European ancestry to enable the bidirectional MR study of T2D and HF followed by a GWAS of diabetes-related HF with collider bias correction (Figure 1). The primary study population consisted of 106,321 MVP participants with T2D diagnosis out of 434,089 with European ancestry, predominantly male. In the GWAS of diabetes-related HF, we included 31,346 HF individuals with comorbid T2D and 74,975 non-HF diabetes controls (Table 1). Individuals with HF were older and had higher prevalence of obesity, atrial fibrillation, coronary artery disease, chronic kidney disease, hyperlipidemia and hypertension with or without T2D (Table 1). The prevalence of all-cause HF was higher among T2D individuals (29.5%) than that among non-diabetes participants (11.2%). In the UKB, we included 26,431 T2D individuals with European ancestry. Among them, 3,506 developed HF using clinical diagnosis codes (Table S1). Similarly, individuals with HF had significant ($p < 0.001$) older age, higher prevalence of cardiometabolic risk factors, and more comorbidities than the control populations without HF.

Genome-wide meta-analysis of all-cause HF

A total of 10,835,443 SNPs with MAF >1% in any one of the three studies (i.e., MVP, HERMES, and FinnGen) were included in the meta-analysis of all-cause HF among European ancestry. We identified a total of 61 independent GWS loci (Table S2) associated with all-cause HF, including 24 novel loci (Table 2) compared with previous reported HF GWAS.[12–14] Overlapping with a T2D GWAS,[42] 59 out of 61 HF-associated SNPs also had summary statistics and were used as the GIVs for all-cause HF in the two-sample MR analysis (Table S2).

The gene-based test based on the HF GWAS meta-analysis mapped to 19,051 protein coding genes and resulted in 86 statistically significant genes at $p < 2.62 \times 10^{-6}$ (Table S3). Tissue expression analysis revealed several tissue types relevant to the heart and blood vessel, including coronary artery, tibial artery, atrial appendage, left ventricle, and skeletal muscle (Tables S4 and S5). Gene-set analysis showed various protein-protein interaction networks (Table S6), and colocalization analysis provided evidence, suggesting that both the gene expression and most signals related to HF share a common causal variant (Table S7). HF loci overlap with enhancer regions that play a role in controlling gene expression (Table S8), and eQTL analysis identified SNPs that have an impact on the regulation of gene expression in HF-related tissues (Table S9).

Bidirectional MR analysis between T2D and all-cause HF

A total of 82 GIVs for T2D (Table S10) and 59 for HF had summary statistics in both T2D and HF GWAS. IVW-MR method showed significant MR association in both directions (Figure 2; Table S11), suggesting potential causal effect of T2D on HF (OR 1.07, 95% confidence interval [CI] 1.04–1.10, $p = 7.02 \times 10^{-7}$), as well as potential causal

Please cite this article in press as: Sun et al., Identification and correction for collider bias in a genome-wide association study of diabetes-related heart failure, The American Journal of Human Genetics (2024), https://doi.org/10.1016/j.ajhg.2024.05.018

Table 2. Twenty-four novel genome-wide significant loci associated with all-cause HF

rsID	Gene	Chr.	Pos. (hg19)	EA	NEA	EAF	OR (95% CI)	*p* value
rs28416760	*INPP5B*	1	38409112	T	A	0.73	1.03 (1.02, 1.04)	1.40×10^{-8}
rs17163313	*MIA3*	1	222799625	G	T	0.71	1.03 (1.02, 1.04)	3.54×10^{-8}
rs7564469	*ZEB2*	2	145258445	C	T	0.16	1.04 (1.03, 1.05)	2.57×10^{-9}
rs3820888	*SPATS2L*	2	201180023	C	T	0.40	1.03 (1.02, 1.04)	1.43×10^{-10}
rs6796042	*FOXP1*	3	71530120	A	G	0.62	1.03 (1.02, 1.04)	7.35×10^{-9}
rs17253722	*SHROOM3*	4	77367287	G	A	0.57	1.03 (1.02, 1.04)	4.57×10^{-8}
rs6842241	*EDNRA*	4	148400819	A	C	0.14	1.04 (1.02, 1.05)	4.67×10^{-8}
rs72810976	*CPEB4*	5	173309057	G	A	0.68	1.03 (1.02, 1.04)	1.05×10^{-8}
rs117321970	*FHL5*	6	97071980	T	C	0.05	1.07 (1.05, 1.10)	2.81×10^{-8}
rs3918226	*NOS3*	7	150690176	T	C	0.08	1.05 (1.03, 1.07)	3.54×10^{-8}
rs4733328	*NRG1*	8	32259246	G	A	0.14	1.04 (1.03, 1.05)	3.44×10^{-8}
rs11774829	*RP11-127H5.1*	8	105978368	T	A	0.88	1.05 (1.03, 1.07)	3.76×10^{-10}
rs7873569	*TMEM245*	9	111796753	A	T	0.57	1.03 (1.02, 1.04)	1.61×10^{-8}
rs71311904	*BDNF*	11	27742447	C	CCATTT	0.82	1.05 (1.03, 1.06)	3.32×10^{-9}
rs113104597	*CHD4*	12	6703172	C	T	0.16	1.04 (1.03, 1.05)	1.21×10^{-8}
rs34682944	*DIP2B*	12	50982864	A	G	0.31	1.03 (1.02, 1.05)	1.07×10^{-8}
rs112403212	*SCARB1*	12	125303254	T	C	0.14	1.05 (1.03, 1.06)	7.99×10^{-9}
rs10161594	*ATP4B*	13	114306243	G	C	0.14	1.04 (1.03, 1.06)	1.09×10^{-8}
rs58472533	*AMN*	14	103385634	G	A	0.20	1.04 (1.03, 1.05)	5.49×10^{-10}
rs17483686	*IREB2*	15	78733390	T	A	0.33	1.03 (1.02, 1.04)	1.55×10^{-9}
rs11634851	*ABHD17C*	15	81028965	G	C	0.45	1.03 (1.02, 1.04)	4.87×10^{-8}
rs11861290	*CMIP*	16	81548522	A	G	0.76	1.04 (1.03, 1.05)	2.27×10^{-10}
rs11656489	*ADORA2B*	17	15837141	G	C	0.19	1.04 (1.02, 1.05)	6.15×10^{-9}
rs17608766	*GOSR2*	17	45013271	C	T	0.15	1.04 (1.03, 1.06)	6.24×10^{-10}

Gene, gene abbreviation of the gene closest to the sentinel SNP; Chr., chromosome; Pos., position; T2D, type 2 diabetes; HF, heart failure; EA, effect allele; NEA, non-effect allele; EAF, effect allele frequency; OR: odds ratio; CI, confidence interval.
Novel locus: a 1 Mb region around the sentinel SNP (±500 kb) not overlapping with any previously reported genome-wide significant locus (±500 kb region centered around the sentinel SNP of each locus). Only the summary statistics of the sentinel SNPs are reported in the table.

effect of HF on T2D (OR 1.60, 95% CI 1.36–1.88, $p = 1.55 \times 10^{-8}$). The MR-Egger method didn't support a significant intercept, which indicates limited pleiotropy. Therefore, we used IVW results as the primary MR estimates in the bidirectional MR analysis. Only the HF effect on T2D showed significant positive association in MR-Egger analysis. After removing 5 (rs10965223, rs635634, rs7903146, rs1061810, and rs1558902) and 2 (rs600038 and rs11642015) outliers for T2D and HF, respectively, the MR-PRESSO analysis showed similar significant MR associations between T2D and HF in both directions using IVM and MR-Egger methods (Figure 2; Table S11).

Using LHC-MR method, we also identified bi-directional relationship between all-cause HF and T2D. Similar to two-sample MR results, T2D is associated with higher risk for HF with moderate effect size (OR 1.09, 95% CI 1.06–1.13, p value 3.95×10^{-7}). Meanwhile, HF is associated with higher risk for T2D with much larger effect size (OR 1.95, 95% CI 1.55–2.44, p value 6.82×10^{-9}).

GWAS of diabetes-related HF with Slope-Hunter correction

In the GWAS of diabetes-related HF among 106,321 individuals with diabetes (31,346 individuals with all-cause HF, 29.5%), we identified nine suggestively significant (p value $<10^{-6}$) loci including three GWS (p value $<5 \times 10^{-8}$) loci associated with diabetes-related HF (Table 3; Figure 3). The inflation factor of the GWAS is 1.04. One diabetes-related HF-associated locus located on chromosome 10 (*TCF7L2*) is strongly associated with T2D but not associated with all-cause HF in the meta-analysis (p value of 0.57), which can be affected by collider bias. After Slope-Hunter correction, the *TCF7L2* locus was no longer associated with diabetes-related HF (OR 1.02, 95% CI 0.99–1.05, p value 0.15). After IER correction, the association between

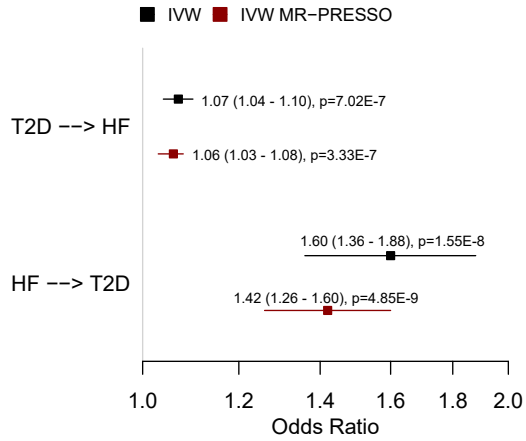

Figure 2. Forest plot of bidirectional MR between T2D and all-cause HF

95% CI of OR is included in the parentheses. IVW, inverse-variance weighted; MR, Mendelian randomization; p, *p* value.

the *TCF7L2* locus remained GWS. However, the associations of *TCF7L2* with diabetes-related HF, or diabetes-adjusted HF diminished after collider bias correction using Slope-Hunter (Tables 3 and S12) and mtCOJO (Table S12). Meanwhile, the other two loci (Figure S1) on chromosome 4 (sentinel SNP rs17513625 close to *PITX2*, OR 1.25, 95% CI 1.16–1.35, *p* value 9.98×10^{-9}) and 9 (sentinel SNP rs4977575 close to *CDKN2B−AS1*, OR 1.06, 95% CI 1.04–1.08, *p* value 2.91×10^{-9}) remained GWS after Slope-Hunter correction for collider bias (Figure 4). Interestingly, the genetic association of chromosome 4 locus with HF was much weaker among 327,768 MVP participants without T2D (OR 1.07, 95% CI 1.00–1.14, *p* value 0.039), presenting an example that T2D may increase the genetic association of HF (interaction *p* value 0.016). We pursued replication of two GWS loci on chromosome 4 and 9 using the UK Biobank study participants with European ancestry (Table S1). After applying the Slope-Hunter correction to the GWAS of diabetes-related HF adjusted for age, sex, and top ten PCs, consistent associations were identified for rs17513625 (*PITX2* locus, OR 1.19, 95% CI 1.02–1.40, *p* value 0.027) and rs4977575 (*CDKN2B−AS1* locus, OR 1.08, 1.03–1.14, *p* value 0.0034), respectively.

Using the *p* value cutoff of 0.001, 10^{-4}, and 10^{-5}, we investigated if the selection of GIVs and the slope estimates are sensitive to the parameter setting in the Slope-Hunter method (Table S13). The estimated slope ranged from −0.198 to −0.219 with overlapping 95% CIs. Different *p* value cutoffs had little impact on Slope-Hunter corrected GWAS of diabetes-related HF. Across all threshold levels, the two GWS loci remained the same, and the *TCF7L2* locus was not significantly associated with diabetes-related HF (*p* value >0.05).

Discussion

The present study aimed to elucidate the relationship between T2D and HF and identify the genetic loci of diabetes-related HF. Using GWS loci from a large meta-analysis of all-cause HF, we conducted a bidirectional MR analysis to investigate the relationship between T2D and HF. The estimates from the two-sample MR strongly supported that not only is T2D a risk factor of HF, but also, HF increases the risk for T2D. As a result, a diabetes-stratified or a diabetes-adjusted HF GWAS may identify spurious genetics associations due to collider bias (both HF and genetic factors can affect T2D). We adopted a recently developed method, Slope-Hunter, to correct for such collider bias in the GWAS of diabetes-related GWAS among over 100,000 individuals with diabetes from the MVP. The Slope-Hunter method assumes that the model-based clustering algorithm correctly identifies the valid GIVs. This tends to be the case when the largest number of similar ratios β^c_{GY}/β_{GX} comes from the valid GIVs.[50] In many simulation scenarios, Slope-Hunter performs well with correct type-1 error and increased power over instrument effect regression. However, Slope-Hunter has poor performance when the invalid GIVs explain more or equal variation in the index event than the valid GIVs and have strong negative correlation of effects.[45] After removing the T2D-associated *TCF7L2* locus by Slope-Hunter correction, we identified two GWS loci associated with diabetes-related HF located on chromosome 4 (*PITX2*) and chromosome 9 (*CDKN2B−AS1*). Although both loci have been associated with all-cause HF,[12,13] the effect size of the SNP (*PITX2*) was larger among T2D individuals than that among non-T2D participants. In addition, the sentinel SNP rs17513625 is weakly correlated with the established atrial fibrillation-associated *PITX2* locus (LD r^2 of 0.115 with rs17042175 in the European ancestry). By definition of collider bias, we anticipated that the collider bias could also affect the diabetes-adjusted GWAS of all-cause HF. Without Slope-Hunter correction, we identified 22 GWS loci associated with all-cause HF among European ancestry (Table S12; Figure S2). Two loci located on chromosome 1 (*C1orf185*) and chromosome 10 (*TCF7L2*) were not GWS after Slope-Hunter correction. Both loci were significantly associated with T2D (Table S12). One unique assumption of IER is that the collider bias *b* is constant across SNPs and may be estimated by the linear regression of β'_{GY} on β_{GX} across many SNPs.[46] This assumption may be violated as the shared genetic component between T2D and HF can be substantial. Therefore, Slope-Hunter can be effective in the correction of collider bias in the present study because the estimate of bias *b* relies on a subset of SNPs with likely causal effect (Tables 3 and S12).

Observational studies consistently demonstrated that diabetes increases the risk for HF. On the other hand, HF induces metabolic impairment, which leads to higher incidence of T2D among individuals with HF than in comparable general populations. Significant MR associations from the present study supported the bidirectional causal relationship between T2D and all-cause HF. Regardless of the directionality of the effects, adults with both diabetes and HF can have 8.8-fold higher mortality rate than

Table 3. Genomic loci associated with diabetes-related HF (genomic loci with $p < 10^{-6}$) with correction for collider bias using Slope-Hunter and instrument effect regression

							Diabetes-related HF GWAS		T2D GWAS		HF GWAS meta-analysis		Diabetes-related HF GWAS after Slope-Hunter correction		Diabetes-related HF GWAS after instrument effect regression correction	
rsID	**Gene**	**Chr.**	**Position**	**EA**	**NEA**	**EAF**	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***
rs602633	*PSRC1*[a]	1	109821511	G	T	0.78	1.06 (1.04, 1.09)	3.83×10^{-7}	1.00 (0.99, 1.02)	0.518	1.05 (1.04, 1.06)	6.75×10^{-17}	1.06 (1.04, 1.09)	3.11×10^{-7}	1.06 (1.04, 1.09)	7.00×10^{-7}
rs17513625	*PITX2*[a,b]	4	111848270	A	G	0.02	1.24 (1.15, 1.34)	1.72×10^{-8}	1.03 (0.98, 1.08)	0.280	1.11 (1.08, 1.14)	6.63×10^{-14}	1.25 (1.16, 1.35)	9.98×10^{-9}	1.23 (1.14, 1.33)	4.95×10^{-8}
rs55730499	*LPA*[a]	6	161005610	T	C	0.07	1.11 (1.07, 1.15)	1.87×10^{-7}	1.00 (0.97, 1.02)	0.767	1.1 (1.08, 1.12)	1.79×10^{-23}	1.11 (1.06, 1.15)	3.02×10^{-7}	1.11 (1.07, 1.15)	1.91×10^{-7}
rs4977575	*CDKN2B–AS1*[a,b]	9	22124744	G	C	0.5	1.06 (1.04, 1.08)	2.40×10^{-8}	1.02 (1.01, 1.03)	1.74×10^{-3}	1.06 (1.05, 1.07)	1.37×10^{-31}	1.06 (1.04, 1.08)	2.91×10^{-9}	1.05 (1.03, 1.08)	2.50×10^{-7}
rs1837530484	*LINC02881*	10	44738619	CA	C	0.91	1.10 (1.06, 1.14)	5.68×10^{-7}	1.00 (0.97, 1.02)	0.681	1.02 (1, 1.05)	0.0306	1.09 (1.06, 1.14)	9.37×10^{-7}	1.10 (1.06, 1.14)	5.30×10^{-7}
rs201426892	*AGAP5*	10	75439094	G	A	0.99	1.78 (1.42, 2.24)	5.17×10^{-7}	1.07 (0.94, 1.21)	0.287	–	–	1.81 (1.44, 2.27)	3.23×10^{-7}	1.76 (1.4, 2.21)	1.10×10^{-6}
rs11196211	*TCF7L2*[c]	10	114817009	A	C	0.69	1.07 (1.05, 1.1)	1.16×10^{-9}	0.80 (0.79, 0.81)	2.50×10^{-235}	1 (0.99, 1.01)	0.569	1.02 (0.99, 1.05)	0.154	1.12 (1.1, 1.15)	8.70×10^{-24}
rs4403799	*AMPD3*	11	10330455	G	A	0.11	1.08 (1.05, 1.12)	9.58×10^{-7}	0.99 (0.97, 1.01)	0.497	1.03 (1.02, 1.05)	5.57×10^{-6}	1.08 (1.05, 1.12)	2.05×10^{-6}	1.08 (1.05, 1.12)	7.55×10^{-7}
rs797765	*SLC6A13*	12	372438	G	A	0.78	1.06 (1.04, 1.09)	8.12×10^{-7}	0.98 (0.96, 0.99)	2.76×10^{-3}	1.02 (1.01, 1.03)	0.0016	1.06 (1.03, 1.08)	7.61×10^{-6}	1.07 (1.04, 1.09)	1.40×10^{-7}

Chr., chromosome; Position, human genome build hg19; EA, effect allele; NEA, non-effect allele; EAF, effect allele frequency; HF, heart failure; GWAS, genome-wide association study; T2D, type 2 diabetes; OR, odds ratio; CI, confidence interval.

[a]GWS loci associated with all-cause HF in the meta-analysis of the European ancestry

[b]GWS association (p value $<5 \times 10^{-8}$) before and after Slope-Hunter correction

[c]GWS association (p value $<5 \times 10^{-8}$) with diabetes-related HF but not significant (p value>0.05) after Slope-Hunter correction.

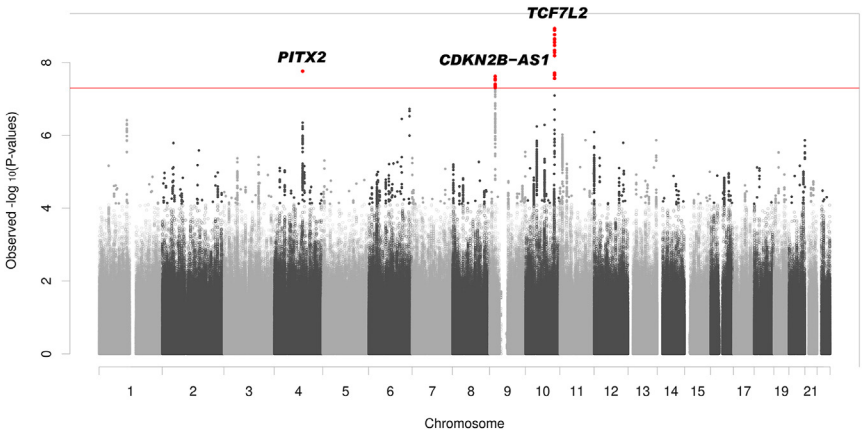

Figure 3. Manhattan plot of diabetes-related HF GWAS without correction for collider bias

Red horizontal line indicates GWS threshold of nominal p value of 5×10^{-8}.

those without HF (32.7 vs. 3.7 per 1,000 person-years).[51] Thus, managing T2D and hyperglycemia can be effective to prevent HF, to mitigate T2D progression, and eventually reduce mortality among individuals with HF. SGLT2 inhibitors are a new class of antidiabetic medications that reduce hyperglycemia through inhibition of glucose reabsorption in the renal proximal tubules. They significantly reduced the risk of HF-related hospitalization and cardiovascular death.[9–11] SGLT2 inhibitors are recommended for individuals with HF irrespective of diabetes status.[52]

Both T2D and HF are complex clinical conditions involving numerous risk factors and pathways. Recent studies identified subtypes of T2D using risk factor and biomarker data that presented differential clinical outcomes.[53,54] Analyses of T2D-associated loci also revealed genetic clusters linking with pathophysiological pathways underlying T2D,[55] supporting the heterogeneity of T2D mechanism. On the other hand, the heterogeneity of HF has been well documented, even among the major clinical subtypes. Based on the measurement of LVEF, recent guidelines categorized HF into HFrEF, HFpEF, HF with mildly reduced EF (HFmrEF), and HF with improved EF (HFimpEF), with HFrEF and HFpEF as the dominant forms.[52] Not surprisingly, HFrEF and HFpEF have distinguishable risk profiles, different response to treatments, and contrasting clinical prognosis. Even within HFpEF subtype, the evidence of heterogeneous subtypes has emerged to support precision treatment and prognosis,[56] which holds the promise for mitigating the growing burden of HFpEF in the aging population. A recent large GWAS of HFrEF and HFpEF also highlighted the different genetic architecture between two HF clinical subtypes, and supported the phenotypic heterogeneity of HFpEF.[12] However, the limited number of HF subtype GWAS and identified loci, the power of the bidirectional MR between T2D and HF subtypes, and the GWAS are suboptimal. Particularly, only one HFpEF-associated loci close to *FTO* has been reported. Since the *FTO* locus is highly pleiotropic, it cannot be used as a GIV of HFpEF in the MR analysis. The future GWASs of HF subtypes would provide more GIVs to robustly estimate the relationship between T2D and HF subtypes and accurately identify genetic loci of HF with comorbid conditions with correction of potential collider bias.

We acknowledge that the MVP and the UKB participants were recruited from the health care systems and may not fully represent the populations in the clinical system or the general population. However, the genetic associations with cardiometabolic diseases reported in the biobank cohorts are consistent with other cohorts.[12,57,58] Recent research has highlighted the potential influence of selection bias on genetic findings.[59] While the genetic associations with cardiometabolic diseases reported in the biobank cohorts are consistent with other cohorts,[12,57,58] selection bias of such large biobank cohorts can impact genetic association findings, including those identified in the present study. Causal effect of T2D on HF could also be related to clinical diagnosis procedures of T2D and HF. For example, people are more likely to be identified with HF if they have T2D would contribute to the causal association. In the present study, we cannot rule out some contribution from possible increased attention to risk factors, including glucose levels along with blood pressure and lipid levels, in the clinical diagnosis of HF.

Conclusion

Global trend of growing T2D and HF requires improved intervention and prevention strategies for diabetes-related HF, a syndrome with high morbidity and mortality. Exploring the genetic architecture of diabetes-related HF would greatly help understand the mechanism and pathophysiology of the condition as shown in recent GWAS of human diseases. However, the complexity of genetic factors underlying T2D and HF, as well as the relationship between them, created a unique challenge in the identification of true genetic associations with diabetes-related HF. We have demonstrated the evidence supporting the bidirectional relationship between T2D and HF, addressed the impact of collider bias on the GWAS of diabetes-related HF, identified and replicated two genetic loci in the MVP and UK Biobank, two large biobank studies. The study design and analytical workflow can be extended to other studies of diabetic complications, particularly outcomes related to HF and HF subtypes. In light of growing precision medicine studies focusing on certain disease subgroups or individuals with specific comorbid conditions, this case study presented the key considerations of epidemiologic, genetic, and biostatistical evidence and methods for such complex disease research in target populations.

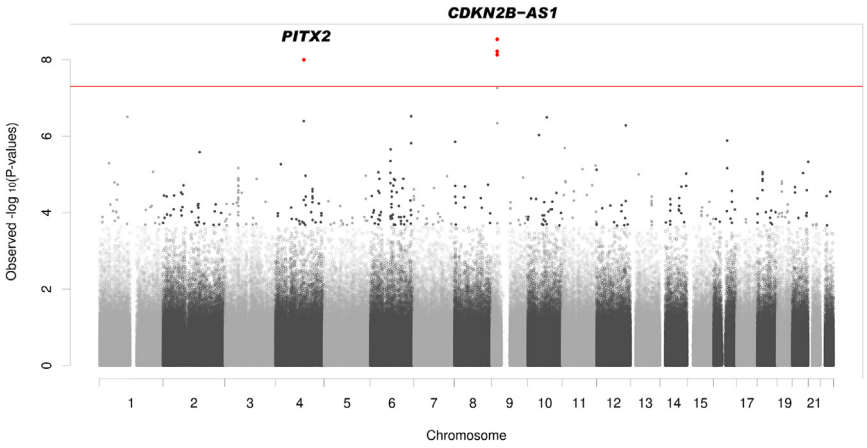

Figure 4. Manhattan plot of diabetes-related HF GWAS after correction for collider bias

Red horizontal line indicates genome-wide significance threshold of nominal p value of 5×10^{-8}. Only independent genetic variants after LD-pruning were included.

Data and code availability

Due to the US Department of Veterans Affairs (VA) regulations and our ethics agreements, the analytic datasets used for this study are not permitted to leave the Million Veteran Program (MVP) research environment and VA firewall. This limitation is consistent with other MVP studies based on VA data. However, the MVP data are made available to researchers with an approved VA and MVP study protocol. The dbGAP accession number for the full summary level association data of the genome-wide association analyses in the MVP and the meta-analysis from this report is dbGAP: phs001672. The only restriction is that use of the data is limited to health/medical/biomedical purposes and does not include the study of population origins or ancestry. Use of the data does include methods development research (e.g., development and testing of software or algorithms), and requestors agree to make the results of studies using the data available to the larger scientific community.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2024.05.018.

Acknowledgments

We are grateful to all the MVP investigators; a list of MVP investigators can be found in the supplemental information. This research has been conducted using the UK Biobank Resource under application number 34031.

This research is supported by funding from the Department of Veterans Affairs Office of Research and Development, Million Veteran Program grants CX001737, BX005831, BX004821, and MVP065 (J.J. and Y.V.S.). This publication does not represent the views of the Department of Veterans Affairs or the United States Government. This research has also been supported in part by National Institutes of Health (NIH) grant P01 HL154996.

Author contributions

Y.V.S. conceptualized the research idea, supervised the study, and wrote the original draft. C.L. and H.Q. analyzed data and contributed to manuscript writing. J.J.Z., J.M.G, P.W.F.W., J.J., and L.S.P. contributed to manuscript review and editing. All authors contributed to discussions about the results and provided feedback on the manuscript.

Declaration of interests

J.J. reports research funding from Amgen, Kowa, Alnylam, Department of Veterans Affairs and National Institutes of Health. Within the past several years, L.S.P. has served on Scientific Advisory Boards for Boehringer Ingelheim and Janssen and has or had research support from Merck, Pfizer, Eli Lilly, Novo Nordisk, Sanofi, PhaseBio, Roche, Abbvie, Vascular Pharmaceuticals, Janssen, Glaxo SmithKline, and the Cystic Fibrosis Foundation and is also a cofounder, officer, board member, and stockholder of a company, Diasyst, Inc., which markets software aimed to help improve diabetes management.

Received: September 14, 2023
Accepted: May 21, 2024
Published: June 18, 2024

Web resources

GTEx, https://gtexportal.org/home/
FinnGEN, https://www.finngen.fi/en/access_results

References

1. Savarese, G., Becher, P.M., Lund, L.H., Seferovic, P., Rosano, G.M.C., and Coats, A.J.S. (2023). Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc. Res. *118*, 3272–3287. https://doi.org/10.1093/cvr/cvac013.
2. Virani, S.S., Alonso, A., Aparicio, H.J., Benjamin, E.J., Bittencourt, M.S., Callaway, C.W., Carson, A.P., Chamberlain, A.M., Cheng, S., Delling, F.N., et al. (2021). Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association. Circulation *143*, e254–e743. https://doi.org/10.1161/CIR.0000000000000950.

3. Heidenreich, P.A., Albert, N.M., Allen, L.A., Bluemke, D.A., Butler, J., Fonarow, G.C., Ikonomidis, J.S., Khavjou, O., Konstam, M.A., Maddox, T.M., et al. (2013). Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ. Heart Fail. *6*, 606–619. https://doi.org/10.1161/HHF.0b013e318291329a.
4. Dunlay, S.M., Givertz, M.M., Aguilar, D., Allen, L.A., Chan, M., Desai, A.S., Deswal, A., Dickson, V.V., Kosiborod, M.N., Lekavich, C.L., et al. (2019). Type 2 Diabetes Mellitus and Heart Failure: A Scientific Statement From the American Heart Association and the Heart Failure Society of America: This statement does not represent an update of the 2017 ACC/AHA/HFSA heart failure guideline update. Circulation *140*, e294–e324. https://doi.org/10.1161/CIR.0000000000000691.
5. Matsushita, K., Blecker, S., Pazin-Filho, A., Bertoni, A., Chang, P.P., Coresh, J., and Selvin, E. (2010). The association of hemoglobin a1c with incident heart failure among people without diabetes: the atherosclerosis risk in communities study. Diabetes *59*, 2020–2026. https://doi.org/10.2337/db10-0165.
6. Held, C., Gerstein, H.C., Yusuf, S., Zhao, F., Hilbrich, L., Anderson, C., Sleight, P., Teo, K.; and ONTARGET/TRANSCEND Investigators (2007). Glucose levels predict hospitalization for congestive heart failure in patients at high cardiovascular risk. Circulation *115*, 1371–1375. https://doi.org/10.1161/CIRCULATIONAHA.106.661405.
7. Triposkiadis, F., Xanthopoulos, A., Bargiota, A., Kitai, T., Katsiki, N., Farmakis, D., Skoularigis, J., Starling, R.C., and Iliodromitis, E. (2021). Diabetes Mellitus and Heart Failure. J. Clin. Med. *10*, 3682. https://doi.org/10.3390/jcm10163682.
8. Kodama, S., Fujihara, K., Horikawa, C., Sato, T., Iwanaga, M., Yamada, T., Kato, K., Watanabe, K., Shimano, H., Izumi, T., and Sone, H. (2020). Diabetes mellitus and risk of new-onset and recurrent heart failure: a systematic review and meta-analysis. ESC Heart Fail. *7*, 2146–2174. https://doi.org/10.1002/ehf2.12782.
9. Zinman, B., Wanner, C., Lachin, J.M., Fitchett, D., Bluhmki, E., Hantel, S., Mattheus, M., Devins, T., Johansen, O.E., Woerle, H.J., et al. (2015). Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. N. Engl. J. Med. *373*, 2117–2128. https://doi.org/10.1056/NEJMoa1504720.
10. Mahaffey, K.W., Neal, B., Perkovic, V., de Zeeuw, D., Fulcher, G., Erondu, N., Shaw, W., Fabbrini, E., Sun, T., Li, Q., et al. (2018). Canagliflozin for Primary and Secondary Prevention of Cardiovascular Events: Results From the CANVAS Program (Canagliflozin Cardiovascular Assessment Study). Circulation *137*, 323–334. https://doi.org/10.1161/CIRCULATIONAHA.117.032038.
11. Wiviott, S.D., Raz, I., Bonaca, M.P., Mosenzon, O., Kato, E.T., Cahn, A., Silverman, M.G., Zelniker, T.A., Kuder, J.F., Murphy, S.A., et al. (2019). Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. N. Engl. J. Med. *380*, 347–357. https://doi.org/10.1056/NEJMoa1812389.
12. Joseph, J., Liu, C., Hui, Q., Aragam, K., Wang, Z., Charest, B., Huffman, J.E., Keaton, J.M., Edwards, T.L., Demissie, S., et al. (2022). Genetic architecture of heart failure with preserved versus reduced ejection fraction. Nat. Commun. *13*, 7753. https://doi.org/10.1038/s41467-022-35323-0.
13. Levin, M.G., Tsao, N.L., Singhal, P., Liu, C., Vy, H.M.T., Paranjpe, I., Backman, J.D., Bellomo, T.R., Bone, W.P., Biddinger, K.J., et al. (2022). Genome-wide association and multi-trait analyses characterize the common genetic architecture of heart failure. Nat. Commun. *13*, 6914. https://doi.org/10.1038/s41467-022-34216-6.
14. Shah, S., Henry, A., Roselli, C., Lin, H., Sveinbjörnsson, G., Fatemifar, G., Hedman, Å.K., Wilk, J.B., Morley, M.P., Chaffin, M.D., et al. (2020). Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. Nat. Commun. *11*, 163. https://doi.org/10.1038/s41467-019-13690-5.
15. Lindgren, M.P., PirouziFard, M., Smith, J.G., Sundquist, J., Sundquist, K., and Zöller, B. (2018). A Swedish Nationwide Adoption Study of the Heritability of Heart Failure. JAMA Cardiol. *3*, 703–710. https://doi.org/10.1001/jamacardio.2018.1919.
16. Kim, J., Jensen, A., Ko, S., Raghavan, S., Phillips, L.S., Hung, A., Sun, Y., Zhou, H., Reaven, P., and Zhou, J.J. (2022). Systematic Heritability and Heritability Enrichment Analysis for Diabetes Complications in UK Biobank and ACCORD Studies. Diabetes *71*, 1137–1148. https://doi.org/10.2337/db21-0839.
17. Zheng, J., Baird, D., Borges, M.C., Bowden, J., Hemani, G., Haycock, P., Evans, D.M., and Smith, G.D. (2017). Recent Developments in Mendelian Randomization Studies. Curr. Epidemiol. Rep. *4*, 330–345. https://doi.org/10.1007/s40471-017-0128-6.
18. Gaziano, J.M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., Whitbourne, S., Deen, J., Shannon, C., Humphries, D., et al. (2016). Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J. Clin. Epidemiol. *70*, 214–223. https://doi.org/10.1016/j.jclinepi.2015.09.016.
19. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209. https://doi.org/10.1038/s41586-018-0579-z.
20. Kurgansky, K.E., Schubert, P., Parker, R., Djousse, L., Riebman, J.B., Gagnon, D.R., and Joseph, J. (2020). Association of pulse rate with outcomes in heart failure with reduced ejection fraction: a retrospective cohort study. BMC Cardiovasc. Disord. *20*, 92. https://doi.org/10.1186/s12872-020-01384-6.
21. Patel, Y.R., Robbins, J.M., Kurgansky, K.E., Imran, T., Orkaby, A.R., McLean, R.R., Ho, Y.L., Cho, K., Michael Gaziano, J., Djousse, L., et al. (2018). Development and validation of a heart failure with preserved ejection fraction cohort using electronic medical records. BMC Cardiovasc. Disord. *18*, 128. https://doi.org/10.1186/s12872-018-0866-5.
22. Rhee, M.K., Ho, Y.L., Raghavan, S., Vassy, J.L., Cho, K., Gagnon, D., Staimez, L.R., Ford, C.N., Wilson, P.W.F., and Phillips, L.S. (2019). Random plasma glucose predicts the diagnosis of diabetes. PLoS One *14*, e0219964. https://doi.org/10.1371/journal.pone.0219964.
23. Aragam, K.G., Chaffin, M., Levinson, R.T., McDermott, G., Choi, S.H., Shoemaker, M.B., Haas, M.E., Weng, L.C., Lindsay, M.E., Smith, J.G., et al. (2019). Phenotypic Refinement of Heart Failure in a National Biobank Facilitates Genetic Discovery. Circulation *139*, 489–501. https://doi.org/10.1161/CIRCULATIONAHA.118.035774.
24. Zhong, H., Magee, M.J., Huang, Y., Hui, Q., Gwinn, M., Gandhi, N.R., and Sun, Y.V. (2020). Evaluation of the Host Genetic Effects of Tuberculosis-Associated Variants Among Patients With Type 1 and Type 2 Diabetes Mellitus. Open Forum Infect. Dis. *7*, ofaa106. https://doi.org/10.1093/ofid/ofaa106.
25. Hunter-Zinck, H., Shi, Y., Li, M., Gorman, B.R., Ji, S.G., Sun, N., Webster, T., Liem, A., Hsieh, P., Devineni, P., et al. (2020). Genotyping Array Design and Data Quality Control in the Million Veteran Program. Am. J. Hum. Genet. *106*, 535–548. https://doi.org/10.1016/j.ajhg.2020.03.004.

26. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873. https://doi.org/10.1093/bioinformatics/btq559.
27. Loh, P.R., Palamara, P.F., and Price, A.L. (2016). Fast and accurate long-range phasing in a UK Biobank cohort. Nat. Genet. *48*, 811–816. https://doi.org/10.1038/ng.3571.
28. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287. https://doi.org/10.1038/ng.3656.
29. Fang, H., Hui, Q., Lynch, J., Honerlaw, J., Assimes, T.L., Huang, J., Vujkovic, M., Damrauer, S.M., Pyarajan, S., Gaziano, J.M., et al. (2019). Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. Am. J. Hum. Genet. *105*, 763–772. https://doi.org/10.1016/j.ajhg.2019.08.012.
30. Kurki, M.I., Karjalainen, J., Palta, P., Sipilä, T.P., Kristiansson, K., Donner, K.M., Reeve, M.P., Laivuori, H., Aavikko, M., Kaunisto, M.A., et al. (2023). FinnGen provides genetic insights from a well-phenotyped isolated population. Nature *613*, 508–518. https://doi.org/10.1038/s41586-022-05473-8.
31. Magi, R., and Morris, A.P. (2010). GWAMA: software for genome-wide association meta-analysis. BMC Bioinf. *11*, 288. https://doi.org/10.1186/1471-2105-11-288.
32. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. *8*, 1826. https://doi.org/10.1038/s41467-017-01261-5.
33. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219. https://doi.org/10.1371/journal.pcbi.1004219.
34. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585. https://doi.org/10.1038/ng.2653.
35. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. Nat. Commun. *6*, 5890. https://doi.org/10.1038/ncomms6890.
36. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. Roy. Stat. Soc. B *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
37. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252. https://doi.org/10.1038/ng.3506.
38. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383. https://doi.org/10.1371/journal.pgen.1004383.
39. Wallace, C. (2020). Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. PLoS Genet. *16*, e1008720. https://doi.org/10.1371/journal.pgen.1008720.
40. Gao, T., He, B., Liu, S., Zhu, H., Tan, K., and Qian, J. (2016). EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. Bioinformatics *32*, 3543–3551. https://doi.org/10.1093/bioinformatics/btw495.
41. Zhang, T., Choi, J., Kovacs, M.A., Shi, J., Xu, M., NISC Comparative Sequencing Program; and Melanoma Meta-Analysis Consortium, Goldstein, A.M., Trower, A.J., Bishop, D.T., et al. (2018). Cell-type-specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes. Genome Res. *28*, 1621–1635. https://doi.org/10.1101/gr.233304.117.
42. Scott, R.A., Scott, L.J., Mägi, R., Marullo, L., Gaulton, K.J., Kaakinen, M., Pervjakova, N., Pers, T.H., Johnson, A.D., Eicher, J.D., et al. (2017). An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes *66*, 2888–2902. https://doi.org/10.2337/db16-1253.
43. Verbanck, M., Chen, C.Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat. Genet. *50*, 693–698. https://doi.org/10.1038/s41588-018-0099-7.
44. Darrous, L., Mounier, N., and Kutalik, Z. (2021). Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics. Nat. Commun. *12*, 7274. https://doi.org/10.1038/s41467-021-26970-w.
45. Mahmoud, O., Dudbridge, F., Davey Smith, G., Munafo, M., and Tilling, K. (2022). A robust method for collider bias correction in conditional genome-wide association studies. Nat. Commun. *13*, 619. https://doi.org/10.1038/s41467-022-28119-9.
46. Dudbridge, F., Allen, R.J., Sheehan, N.A., Schmidt, A.F., Lee, J.C., Jenkins, R.G., Wain, L.V., Hingorani, A.D., and Patel, R.S. (2019). Adjustment for index event bias in genome-wide association studies of subsequent events. Nat. Commun. *10*, 1561. https://doi.org/10.1038/s41467-019-09381-w.
47. Bowden, J., Del Greco M, F., Minelli, C., Davey Smith, G., Sheehan, N.A., and Thompson, J.R. (2016). Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. Int. J. Epidemiol. *45*, 1961–1974. https://doi.org/10.1093/ije/dyw220.
48. Cook, J.R., and Stefanski, L.A. (1994). Simulation-Extrapolation Estimation in Parametric Measurement Error Models. J. Am. Stat. Assoc. *89*, 1314–1328. https://doi.org/10.2307/2290994.
49. Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., Robinson, M.R., McGrath, J.J., Visscher, P.M., Wray, N.R., and Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat. Commun. *9*, 224. https://doi.org/10.1038/s41467-017-02317-2.
50. Cai, S., Hartley, A., Mahmoud, O., Tilling, K., and Dudbridge, F. (2022). Adjusting for collider bias in genetic association studies using instrumental variable methods. Genet. Epidemiol. *46*, 303–316. https://doi.org/10.1002/gepi.22455.
51. Johansson, I., Edner, M., Dahlström, U., Näsman, P., Rydén, L., and Norhammar, A. (2014). Is the prognosis in patients with diabetes and heart failure a matter of unsatisfactory management? An observational study from the Swedish Heart Failure Registry. Eur. J. Heart Fail. *16*, 409–418. https://doi.org/10.1002/ejhf.44.
52. Heidenreich, P.A., Bozkurt, B., Aguilar, D., Allen, L.A., Byun, J.J., Colvin, M.M., Deswal, A., Drazner, M.H., Dunlay, S.M., Evers, L.R., et al. (2022). 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: Executive Summary: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines.

Circulation *145*, e876–e894. https://doi.org/10.1161/CIR.0000000000001062.

53. Ahlqvist, E., Storm, P., Käräjämäki, A., Martinell, M., Dorkhan, M., Carlsson, A., Vikman, P., Prasad, R.B., Aly, D.M., Almgren, P., et al. (2018). Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol. *6*, 361–369. https://doi.org/10.1016/S2213-8587(18)30051-2.
54. Hall, H., Perelman, D., Breschi, A., Limcaoco, P., Kellogg, R., McLaughlin, T., and Snyder, M. (2018). Glucotypes reveal new patterns of glucose dysregulation. PLoS Biol. *16*, e2005143. https://doi.org/10.1371/journal.pbio.2005143.
55. Udler, M.S., Kim, J., von Grotthuss, M., Bonàs-Guarch, S., Cole, J.B., Chiou, J., Christopher D Anderson on behalf of METASTROKE and the ISGC, Boehnke, M., Laakso, M., Atzmon, G., et al. (2018). Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. PLoS Med. *15*, e1002654. https://doi.org/10.1371/journal.pmed.1002654.
56. Shah, S.J., Katz, D.H., Selvaraj, S., Burke, M.A., Yancy, C.W., Gheorghiade, M., Bonow, R.O., Huang, C.C., and Deo, R.C. (2015). Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation *131*, 269–279. https://doi.org/10.1161/CIRCULATIONAHA.114.010637.
57. Vujkovic, M., Keaton, J.M., Lynch, J.A., Miller, D.R., Zhou, J., Tcheandjieu, C., Huffman, J.E., Assimes, T.L., Lorenz, K., Zhu, X., et al. (2020). Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. Nat. Genet. *52*, 680–691. https://doi.org/10.1038/s41588-020-0637-y.
58. Tcheandjieu, C., Zhu, X., Hilliard, A.T., Clarke, S.L., Napolioni, V., Ma, S., Lee, K.M., Fang, H., Chen, F., Lu, Y., et al. (2022). Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. Nat. Med. *28*, 1679–1692. https://doi.org/10.1038/s41591-022-01891-3.
59. Schoeler, T., Speed, D., Porcu, E., Pirastu, N., Pingault, J.B., and Kutalik, Z. (2023). Participation bias in the UK Biobank distorts genetic associations and downstream analyses. Nat. Human Behav. *7*, 1216–1227. https://doi.org/10.1038/s41562-023-01579-9.