Inference of causal relationships between genetic risk factors for cardiometabolic phenotypes and female-specific health conditions

Brenda Xiao
University of Pennsylvania
Digna Velez Edwards
Vanderbilt University Medical Center
Anastasia Lucas
University of Pennsylvania
Theodore Drivas
University of Pennsylvania
Kathryn Gray
Brigham and Women's Hospital
Brendan Keating
University of Pennsylvania
Chunhua Weng
Columbia University
Gail Jarvik
University of Washington https://orcid.org/0000-0002-6710-8708
Hakon Hakonarson
Center for Applied Genomics, Children's Hospital of Philadelphia, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvana
Leah Kottyan
Cincinnati Children's Hospital Medical Center https://orcid.org/0000-0003-3979-2220
Noémie Elhadad
Columbia University
Wei-Qi Wei
Vanderbilt University Medical Center https://orcid.org/0000-0003-4985-056X
Yuan Luo
Northwestern University https://orcid.org/0000-0003-0195-7456
Dokyoon Kim
University of Pennsylvania https://orcid.org/0000-0002-4592-9564
Marylyn Ritchie
University of Pennsylvania
Shefali Verma (Shefali.SetiaVerma@pennmedicine.upenn.edu)
University of Pennsylvania https://orcid.org/0000-0001-5216-4670

Article

Keywords:

Posted Date: March 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1386628/v1

Abstract

Cardiometabolic diseases are highly comorbid, but their relationship with female-specific health conditions (breast cancer, endometriosis, pregnancy complications) is understudied. Using electronic health record data from 71,008 ancestrally diverse females, we examined relationships between 23 obstetrical/gynecological conditions and 4 cardiometabolic phenotypes (body mass index [BMI], coronary artery disease [CAD], type 2 diabetes [T2D], and hypertension) using cross-trait genetic correlation analyses, polygenic risk scores (PRSs), Mendelian randomization (MR), and chronology analyses. We observed 29 significant associations between cardiometabolic PRSs and obstetrical/gynecological conditions (BMI and endometrial cancer; BMI and polycystic ovarian syndrome [PCOS)]; T2D and gestational diabetes; T2D and PCOS). MR analysis provided additional evidence of independent causal effects. We also identified an inverse association between CAD and breast cancer. High cardiometabolic PRSs were associated with early development of PCOS and gestational hypertension. We conclude that polygenic susceptibility to cardiometabolic traits is associated with elevated risk of certain female-specific health conditions.

Introduction

Cardiometabolic diseases such as coronary artery disease (CAD), obesity, hypertension, and type 2 diabetes (T2D) are among the leading causes of death in the world and are highly comorbid[5,6,7]. Many studies have shown that the pathophysiology of cardiometabolic diseases affects males and females differently[8]. Although cardiometabolic diseases have many sequelae that affect both sexes, including depression, anxiety, chronic obstructive pulmonary disease (COPD), and cancer[9,10], they disproportionately affect females because of their links to female-specific health conditions, such as preeclampsia, gestational diabetes, stillbirth, and pregnancy loss[11,12]. Multiple lines of evidence link female health conditions to cardiometabolic diseases. Individuals who develop preeclampsia during pregnancy are more likely to develop cardiovascular diseases and hypertension after pregnancy[13,14]. Obesity is closely linked to polycystic ovarian syndrome (PCOS), and individuals with PCOS are at high risk of developing T2D[15–17]. Individuals with endometriosis are at high risk of developing various cancers and cardiovascular diseases such as myocardial infarction and ischemic heart disease[18,19]. Despite this evidence, the relationship between female-specific health conditions and cardiometabolic phenotypes, and particularly the potential for shared genetic risk, is understudied. Investigation of shared genetic risks between cardiometabolic diseases and female-specific health conditions has the potential to reveal genetic cluster-based risk factors that can be used to improve screening practices for many diseases in high-risk patients.

Genome-wide association studies (GWAS) have exposed common genetic etiologies among diseases such as T2D, obesity, sleep apnea, hypertension, Alzheimer's disease, and many types of cancer[1–4]. There is evidence of impacts on cardiometabolic phenotypes for more than 1,000 different loci, but the effect size of any single variant is generally minimal. To more accurately estimate an individual's overall risk of disease, researchers have used GWAS results to calculate polygenic risk scores (PRSs), which sum the effects of common single-nucleotide polymorphisms (SNPs) on a given phenotype[25] and can be used to predict the risks of many cardiometabolic diseases and comorbidities[26,27].

To investigate the effects of genetic risk of cardiometabolic phenotypes on female health conditions, we measured PRSs for cardiometabolic phenotypes and then determined their correlations with female-specific health conditions documented in electronic health records (EHRs). Prior studies successfully used PRSs as genetic risk factors linking multiple adverse phenotypes[29,30]. PRS-based association tests have the advantage that they are based on unvarying risk factors (i.e., inherited genetic burden), and they make fewer assumptions than association tests based on individual genetic factors. We hypothesized that the genetic risks of adverse cardiometabolic phenotypes can explain a significant portion of phenotypic variance in female-specific health conditions. To test this hypothesis, we obtained genotypic and

phenotypic data on a wide range of female-specific health conditions from the Penn Medicine BioBank (PMBB) and the Electronic Medical Records and Genomics (eMERGE) network. We first measured pairwise genetic correlations between cardiometabolic phenotypes and female-specific health conditions to determine whether there was a common genetic basis. We then estimated the significance of associations between the genetic risks of adverse cardiometabolic phenotypes and female-specific health conditions. In addition, we evaluated potential causal relationships between genetic risk scores for cardiometabolic phenotypes and female-specific health conditions using Mendelian randomization (MR) approaches. Furthermore, because EHR data provide an opportunity to map disease prevalence by age, we generated a chronological map of female-specific health conditions in individuals with high and low genetic risk scores to understand age-specific prevalence in high-risk and low-risk individuals.

Methods

The authors have individual-level access to genotype data and medical records from the PMBB and the eMERGE network datasets.

Study populations

Penn Medicine BioBank

The PMBB is a University of Pennsylvania academic biobank that recruits patient participants from the University of Pennsylvania Health System surrounding the greater Philadelphia area in the United States. PMBB links patient genotype data with detailed EHR information. Currently, PMBB consists of genome-wide and whole-exome sequence data of approximately 45,000 samples. The PMBB cohort is diverse, with over 25% of its participants being of African ancestry. PMBB genotype data were imputed to the TOPMED Reference panel using the Michigan Imputation server. We included 21,837 female participants from PMBB in this study (Table 1). The stratified analyses in this study included only participants of European or African ancestry, while excluding those of Asian or Hispanic ancestry because of sample size limitations.

Table 1
Sample sizes in the PMBB and eMERGE datasets overall and stratified by ancestry.

Phenotype	PMBB Sample Size (N Cases)			eMERGE Sample Size (N Cases)		
	All	European	African	All	European	African
Cardiometabolic Phenotypes						
BMI	20,209	12,344	6,744	39,403	32,880	5,555
CAD	21,837 (3,002)	13,515 (1,956)	7,039 (955)	49,171 (9,597)	38,918 (7,953)	9,067 (1,515)
DBP	21,612	13,343	6,994	NA	NA	NA
Hypertension	21,837 (10,278)	13,515 (5,640)	7,039 (4,286)	49,171 (27,685)	38,918 (21,883)	9,067 (5,265)
T2D	21,837 (4,388)	13,515 (1,969)	7,039 (2,221)	49,171 (12,403)	38,918 (8,999)	9,067 (3,090)
Female Health Phenotypes						
Breast cancer	21,837 (1,621)	13,515 (1,621)	7,039 (415)	49,171 (4,148)	38,918 (3,639)	9,067 (443)
Cervical cancer	21,837 (105	13,515 (53)	7,039 (50)	49,171 (332)	38,918 (263)	9,067 (60)
Ectopic pregnancy	2,808 (1,779)	1,201 (827)	1,319 (757)	3,078 (823)	1,975 (570)	641 (172)
Endometrial cancer	21,837 (286)	13,515 (183)	7,039 (90)	49,171 (771)	38,918 (680)	9,067 (83)
Endometriosis	21,837 (701)	13,515 (320)	7,039 (340)	49,171 (2,314)	38,918 (1,891)	9,067 (354)
Excessive fetal growth	693 (85)	293 (37)	329 (42)	2,433 (627)	1,618 (422)	437 (162)
Gestational diabetes	2,655 (523)	1,111 (193)	1,262 (248)	3,174 (762)	2,005 (445)	719 (251)
Gestational hypertension	2,666 (631)	1,118 (256)	1,275 (324)	3,135 (703)	1,980 (411)	711 (253)
Intrauterine death	685 (57)	289 (23)	324 (27)	2,156 (64)	1,438 (44)	358 (19)

Phenotype	PMBB Sample Size (N Cases)			eMERGE Sample Size (N Cases)		
	All	European	African	All	European	African
Miscarriage	2,934 (389)	1,273 (199)	1,360 (151)	3,568 (823)	2,323 (577)	740 (200)
Ovarian cancer	21,837 (305)	13,515 (191)	7,039 (89)	49,171 (1,589)	38,918 (1,359)	9,067 (197)
Placenta abruption/previa	1,192 (396)	482 (157)	586 (195)	2,674 (997)	1,697 (672)	583 (262)
Polycystic ovarian syndrome	21,837 (736)	13,515 (387)	7,039 (272)	49,171 (1,006)	38,918 (750)	9,067 (209)
Poor fetal growth	792 (202)	325 (78)	388 (107)	2,209 (167)	1,478 (115)	360 (27)
Postpartum depression	924 (384)	374 (136)	466 (226)	2,417 (555)	1,579 (349)	457 (177)
Postpartum hemorrhage	846 (283)	350 (108)	408 (146)	2,320 (401)	1,544 (261)	408 (109)
Preeclampsia	2,631 (452)	1,100 (149)	1,264 (272)	3,132 (702)	1,974 (406)	713 (249)
Preterm birth	687 (66)	284 (21)	332 (40)	2,144 (57)	1,433 (35)	350 (16)
Stillbirth	649 (17)	270 (3)	311 (12)	2,123 (8)	1,417 (7)	347 (0)
Uterine cancer	21,837 (113)	13,515 (66)	7,039 (43)	49,171 (418)	38,918 (348)	9,067 (64)
Uterine fibroid	21,837 (1,570)	13,515 (516)	7,039 (984)	49,171 (5,711)	38,918 (4,103)	9,067 (1,455)
Vaginal cancer	21,837 (23)	13,515 (13)	7,039 (9)	49,171 (109)	38,918 (89)	9,067 (18)
Vulvar cancer	21,837 (41)	13,515 (25)	7,039 (14)	49,171 (120)	38,918 (97)	9,067 (21)

Electronic Medical Records and Genomics

The eMERGE network is a nationwide consortium that combines genome-wide sequence data with EHRs from several health systems across the United States, with most participants coming from the Geisinger Health System and Vanderbilt University. eMERGE data were imputed to the Haplotype Reference Consortium panel using the Michigan Imputation Server. Like the PMBB cohort, the eMERGE cohort is diverse across ancestries (~ 20% samples of non-European ancestry) and ages. This study included data from 49,171 female patients in the eMERGE network born after 2001(Table 1). Low sample sizes of participants of Asian and Hispanic ancestry also limited ancestry-stratified analyses to include only those of European and African ancestry.

GWAS of cardiometabolic phenotypes

To determine genetic correlations and calculate PRSs, we collected quantitative data on the effects of genetic variants on cardiometabolic phenotypes from several large GWAS. Given the diverse nature of our study population, we obtained the largest publicly available multi-ancestry GWAS summary statistics for four cardiometabolic phenotypes: obesity (measured as body mass index [BMI]), CAD, hypertension (measured as diastolic blood pressure [DBP], systolic blood pressure [SBP], and pulse pressure [PP]), and T2D. Summary statistics and source studies for each phenotype are shown in Table 2.

Table 2
GWAS datasets used to calculate genetic correlations and polygenic risk scores and the number of SNPs used from each GWAS to calculate the corresponding PRS.

Phenotype	Ancestries Included	Source	Sample Size (N Cases)	PMID	N SNPs included in PRS
Type 2 Diabetes	EUR, AFR, EAS, SAS, HIS	Vujkovic et al., Nat Gen, 2020, Million Veterans Program	1,407,282 (228,499)	32541925	PMBB: 1,023,697 eMERGE: 716,330
Body Mass Index	EUR, AFR, EAS, SAS, HIS	Justice et al., Nat Com, 2017, GIANT	241,258	28443625	PMBB: 885,143 eMERGE: 614,668
Hypertension (DBP, SBP, PP)	EUR, AFR, EAS, SAS, HIS	Giri et al., Nat Gen, 2019, Million Veterans Program	318,891	30578418	PMBB: 1,024,567 eMERGE: 715,471
Coronary Artery Disease	EUR, EAS, SAS, HIS, AFR	van der Harst et al., Circ Res, 2018, CARDIoGRAMplusC4D + UKBB	547,261 (122,733)	29212778	PMBB: 981,480 eMERGE: 681,029

**DBP = Diastolic Blood Pressure; SBP = Systolic Blood Pressure; PP = Pulse Pressure; EUR = European ancestry; EAS = East Asian ancestry; SAS = South Asian ancestry; HIS = Hispanic ancestry; AFR = African ancestry; MVP = Million Veterans Program; GIANT = The Genetic Investigation of ANthropometric Traits; UKB = UK BioBank

Genome-wide associations of female-specific health conditions

Large multi-ancestry GWAS are not available for most of the female-specific health conditions evaluated in this study. Therefore, we used PLINK version 1.90 to conduct GWAS of female-specific health conditions documented in the PMBB and eMERGE datasets[31]. We filtered the genetic variants in the PMBB and eMERGE imputed datasets to include only

variants with imputation quality $R^2 > 0.3$ and minor allele frequency > 0.01. We then conducted a meta-analysis of the GWAS of the two cohorts using PLINK.

Genetic correlation calculation

We calculated pairwise genetic correlations between cardiometabolic phenotypes and female-specific health conditions using linkage disequilibrium score regression (LDSC)[32]. LDSC accounts for linkage disequilibrium (LD) among SNPs by using an external reference panel that should match the ancestry distribution of the corresponding GWAS. We generated a multi-ancestry LD reference panel using the HapMap3 SNPs (~ 1 M common variants) from the entire 1000 Genomes population[33].

Polygenic risk scores

PRSs calculated using a GWAS of one particular ancestry group tend to perform poorly when applied to individuals from a different ancestry group[25,34,35]. To accurately calculate PRSs for our diverse cohorts, we calculated PRSs for the cardiometabolic phenotypes using the same publicly available GWAS used in the genetic correlation analyses. Weights for each SNP were calculated using PRS-CS (version from Apr 24, 2020)[37]. Like LDSC, PRS-CS requires a reference panel that matches the ancestry distribution of the target dataset. We generated a multi-ancestry LD reference panel using the HapMap SNPs from the 1000 Genomes populations. We then used PLINK (v1.90) to identify LD blocks and calculate the LD between the SNPs in each block. For PRS-CS, the global shrinkage parameter phi was fixed to 0.01, and default values were selected for all other parameters. PRSs were then calculated using the weights with PLINK. Only the SNPs in the target dataset, summary statistics, and LD reference panel were included in the PRSs. The numbers of SNPs used for each PRS calculation are listed in Table 2. The scores were then normalized to obtain meaningful beta coefficients.

To evaluate the power of the PRSs, we tested their performance in prediction of the corresponding primary phenotypes in the summary statistics. We could not obtain quantitative measurements of blood pressure for participants in the eMERGE cohort, and PP and SBP measurements were not curated in the PMBB cohort. Therefore, we evaluated the performance of the PRSs for blood pressure traits (SP, DBP, PP) in the eMERGE cohort based on hypertension case-control phenotypes, and we evaluated the performance of the PRS blood pressure traits in the PMBB cohort based on DBP or hypertension (for SBP and PP) as outcomes. We constructed logistic regression models for binary phenotypes (CAD, T2D, and hypertension) and evaluated PRS performance by the area under the receiver-operator curves (AUCs) using the *pROC* package in R. Similarly, we constructed linear regression models for continuous phenotypes (BMI and DBP) and evaluated them by the R^2 using the *glm* function in R. The regression models used birth year and the first five principal components (PCs) as covariates. We tested PRS performance overall as well as for subgroups of European or African ancestry.

Phenotype data

Cases and controls for each phenotype were defined using International Classification of Disease (ICD) diagnosis codes. Participants were coded as cases of a given phenotype if their records contained at least one of the corresponding ICD codes. For pregnancy-related phenotypes, participants were only considered controls if their records had at least one pregnancy-related ICD code and no ICD codes for relevant complications (such as miscarriage) during pregnancy. Participants were counted as controls for cardiometabolic phenotypes and all non-pregnancy related health conditions if their records did not contain any relevant ICD code. The complete list of ICD codes used to include or exclude participants as cases and controls can be found in **Supplementary Table 1**. Using these definitions, we determined the sample size for each phenotype in the eMERGE and PMBB cohorts (Table 1).

PRS association analyses

We tested the association between each cardiometabolic PRS and female-specific health condition by fitting separate logistic regression models adjusted by birth year and the first five PCs. We conducted this analysis for all participants and for subsets of participants of European or African ancestry. To account for biases from multiple hypothesis testing, we determined if associations passed a false discovery rate significance threshold of 0.05, adjusting the p values with the number of female-specific health conditions (23) tested. The logistic regressions were performed using the *glm* function in R. The results from the PMBB and eMERGE cohorts were combined using the *rma* function from the *metafor* R package under the restricted maximum-likelihood estimator model[36]. We used PheWAS-View to visualize the results[38]. We then created prevalence plots for each significant association. The participants were divided into quintiles based on the PRS for each condition, and the percentage of cases was calculated for each quintile.

Mendelian randomization

To identify evidence of causality between cardiometabolic phenotypes and female-specific health conditions, we performed one-sample MR for 29 significant associations from the PRS analyses using the *ivreg* function in the *ivpack* R package. Cardiometabolic PRSs were used as genetic instruments, with birth year and the first five PCs included as covariates. The results were then combined in a meta-analysis using the *metafor* R package. Because the small sample size for some of the health conditions could limit the power of the analysis, we also performed two-sample MR for the same associations using the inverse variance weighted (IVW) method in the *twoSampleMR* package in R[39]. MR sensitivity analyses were conducted using the weighted median and MR Egger methods in the same package. The genetic instruments for the cardiometabolic phenotypes in the two-sample MR were the significant SNPs in the GWAS summary statistics used to calculate the PRSs. The SNPs were pruned according to LD patterns among the 1000 Genomes HapMap SNPs, and the remaining representative SNPs were included in the analysis. Matching genetic instruments for the female-specific health conditions were obtained from publicly available GWAS through the *twoSampleMR* package (**Supplementary Table 2**). Since these GWAS for female-specific health conditions were conducted in only European ancestry populations, we performed two-sample MR for only associations that were significant in all participants or participants of European ancestry.

Chronology analyses

We divided the participants into high-risk and low-risk groups according to each cardiometabolic PRS. High PRS was defined as the top quintile (> 80th percentile), and low PRS was defined as the bottom quintile (< 20th percentile). We considered age at the first occurrence of each female-specific health condition from the ICD records. For pregnancy-related conditions, the participants were split into three age groups: <25, 25–40, and 40–55 years. We excluded participants who were over 55 years of age at the first occurrence of a pregnancy-related condition because of low sample sizes and potential errors in diagnosis coding. For all other conditions, the participants were split into five age groups: <25, 25–40, 40–55, 55–70, and > 70 years. We then examined the combined case prevalence of each health condition within the high and low PRS groups in each cohort across all age groups.

Results

Genetic correlations among cardiometabolic phenotypes and female-specific health conditions

We found thirteen statistically significant correlations between the cardiometabolic phenotypes and six female-specific health conditions (Fig. 1A). BMI was correlated with breast cancer (R_g=-0.179, p = 0.011) and PCOS (R_g=0.4, p = 0.007); CAD with breast cancer (R_g=-0.199, p = 0.011), PCOS (R_g=0.216, p = 0.04), and postpartum depression (R_g=0.204, p = 0.025); PP with gestational hypertension (R_g=0.481, p = 0.0025); SBP with breast cancer (R_g=-0.153, p = 0.041) and

gestational hypertension (R_g=0.523, p = 0.0017); and T2D with breast cancer (R_g=-0.196, p = 0.0001), excessive fetal growth (R_g=0.126, p = 0.044), gestational diabetes (R_g=0.529, p = 0.011), gestational hypertension (R_g=0.237, p = 0.028), and PCOS (R_g=0.316, p = 0.0093).

PRS performance for predicting primary phenotypes

We calculated a PRS for each cardiometabolic phenotype and confirmed the distributions of the raw and normalized scores (**Supplementary Figs. 1, 2, 3, 4, 5, and 6**). The full model for all PRSs generally performed well in predicting the corresponding primary phenotypes across ancestry groups (Table 3). The covariate-only (null) model performed better for participants of African ancestry than for participants of European ancestry, whereas the PRS-only model performed better for participants of European ancestry than for participants of African ancestry. We calculated the difference between the full and null models for all PRSs in each cohort and found that the PRSs improved the predictive performance more for participants of European ancestry than for participants of African ancestry (p = 0.000488, Wilcoxon signed rank test). Thus, the cardiometabolic PRSs were more accurate in individuals with European ancestry than in individuals with African ancestry, but they improved predictive models in both groups.

Table 3
Effect estimates for tests of association of PRS with primary phenotype in the eMERGE and PMBB datasets

Phenotype	Ancestry	N Total	N Cases	R^2 or AUC of Full Model	Beta	SE	P-value	R^2 or AUC of Null Model	R^2 or AUC of PRS-only Model
eMERGE									
BMI	All	39,338		0.08494	2.378	0.05044	0	0.0334	0.0593
	EUR	32,880		0.07283	2.334	0.05075	0	0.0132	0.0595
	AFR	5,555		0.0697	2.772	0.2457	3.42E-29	0.0485	0.0229
CAD	All	49,171	9,597	0.784	0.3091	0.01318	1.62E-121	0.7745	0.5474
	EUR	38,918	7,953	0.7733	0.3039	0.01391	7.65E-106	0.7624	0.5629
	AFR	9,067	1,515	0.8228	0.3	0.04424	1.23E-11	0.8197	0.5462
DBP (Hypertension)	All	49,171	27,685	0.8109	0.1387	0.01347	7.17E-25	0.8099	0.529
	EUR	38,918	21,883	0.7994	0.1504	0.01459	6.34E-25	0.798	0.5355
	AFR	9,067	5,265	0.8619	0.1709	0.03923	1.32E-05	0.8612	0.5068
PP (Hypertension)	All	49,171	27,685	0.8115	0.1644	0.01344	2.15E-34	0.8099	0.526
	EUR	38,918	21,883	0.8001	0.1673	0.01413	2.23E-32	0.798	0.5276
	AFR	9,067	5,265	0.8613	0.0974	0.04696	0.0381	0.8612	0.5044
SBP (Hypertension)	All	49,171	27,685	0.8131	0.2977	0.01665	1.86E-71	0.8099	0.5352
	EUR	38,918	21,883	0.8024	0.3143	0.01756	1.1E-71	0.798	0.5446
	AFR	9,067	5,265	0.8622	0.3127	0.06055	2.42E-07	0.8612	0.5031
T2D	All	49,171	12,403	0.7193	0.6334	0.01914	3.02E-240	0.691	0.6163
	EUR	38,918	8,999	0.6963	0.6275	0.0202	9.05E-212	0.655	0.6145
	AFR	9,067	3,090	0.7794	0.6444	0.06331	2.49E-24	0.7724	0.5483
PMBB									

**BMI = Body Mass Index; CAD = Coronary artery disease, T2D = Type 2 Diabetes; DBP = Diastolic Blood Pressure; SBP = Systolic Blood Pressure; PP = Pulse Pressure; EUR = European ancestry; AFR = African ancestry; Beta coefficients and standard errors are per SD PRS in the full model. Covariates included: birth year and the first five PCs.

Phenotype	Ancestry	N Total	N Cases	R^2 or AUC of Full Model	Beta	SE	P-value	R^2 or AUC of Null Model	R^2 or AUC of PRS-only Model
BMI	All	20,209		0.1542	2.658	0.08244	2.05E-222	0.1108	0.1417
	EUR	12,344		0.07473	2.621	0.08679	2.66E-193	0.006422	0.06782
	AFR	6,744		0.03203	2.708	0.2089	5.75E-38	0.008037	0.02711
CAD	All	21,837	3,002	0.8235	0.3571	0.02354	5.73E-52	0.8148	0.5619
	EUR	13,515	1,956	0.8303	0.3894	0.02732	4.44E-46	0.8185	0.5929
	AFR	7,039	955	0.799	0.2566	0.05214	8.60E-07	0.7955	0.5389
DBP	All	21,612		0.0425	1.036	0.06661	3.05E-54	0.03183	0.02905
	EUR	13,343		0.01534	0.9964	0.08064	7.01E-35	0.004144	0.01213
	AFR	6,994		0.04559	1.104	0.129	1.35E-17	0.03571	0.01375
PP (Hypertension)	All	21,837	10,278	0.8125	0.1942	0.02328	7.18E-17	0.8112	0.602
	EUR	13,515	5,640	0.7797	0.1918	0.02638	3.53E-13	0.7778	0.5272
	AFR	7,039	4,286	0.8389	0.1886	0.05376	0.00045	0.8383	0.5391
SBP (Hypertension)	All	21,837	10,278	0.8144	0.327	0.02523	2.08E-38	0.8112	0.6097
	EUR	13,515	5,640	0.7813	0.279	0.02846	1.11E-22	0.7778	0.5397
	AFR	7,039	4,286	0.8415	0.4505	0.05923	2.81E-14	0.8383	0.5649
T2D	All	21,837	4,388	0.754	0.8136	0.03523	5.49E-118	0.7287	0.6713
	EUR	13,515	1,969	0.712	0.8418	0.04533	5.55E-77	0.6634	0.6323
	AFR	7,039	2,221	0.7342	0.7896	0.05947	3.11E-40	0.7168	0.5919

**BMI = Body Mass Index; CAD = Coronary artery disease, T2D = Type 2 Diabetes; DBP = Diastolic Blood Pressure; SBP = Systolic Blood Pressure; PP = Pulse Pressure; EUR = European ancestry; AFR = African ancestry; Beta coefficients and standard errors are per SD PRS in the full model. Covariates included: birth year and the first five PCs.

Associations between cardiometabolic risks and female-specific health conditions

We detected numerous associations between cardiometabolic PRSs and female health conditions in the meta-analysis of both cohorts (Fig. 1B, 1C, and 1D). Twenty-nine associations were statistically significant after correction for multiple hypothesis testing (Table 4). In the meta-analysis, for all participants and participants of European ancestry, PRS_{BMI} was associated with endometrial cancer ($beta_{all}$=0.24, se_{all}=0.046, p_{all}=$9.4 \cdot 10^{-8}$; $beta_{eur}$=0.28, se_{eur}=0.083, p_{eur}=0.00075), gestational diabetes ($beta_{all}$=0.23, se_{all}=0.051, p_{all}=$6 \cdot 10^{-6}$; $beta_{eur}$=0.29, se_{eur}=0.061, p_{eur}=$2.6 \cdot 10^{-6}$), and PCOS ($beta_{all}$=0.27, se_{all}=0.039, p_{all}=$2.4 \cdot 10^{-12}$; $beta_{eur}$=0.28, se_{eur}=0.044, p_{eur}=$2.1 \cdot 10^{-10}$). PRS_{BMI} was also associated with breast cancer in all participants ($beta_{all}$=-0.071, se_{all}=0.022, p_{all}=0.0016). These results suggest that an increased genetic risk of obesity heightens the risks of endometrial cancer, gestational diabetes, and PCOS but decreases the risk of breast cancer. PRS_{CAD} was negatively associated with breast cancer for all participants and participants of European ancestry ($beta_{all}$=-0.072, se_{all}=0.015, p_{all}=$1 \cdot 10^{-6}$; $beta_{eur}$=-0.071, se_{all}=0.016, p_{eur}=$5.3 \cdot 10^{-6}$). This finding suggests that individuals, particularly those of European ancestry, with high PRS_{CAD} are at relatively low risk of breast cancer compared with individuals with low PRS_{CAD}. PRS_{T2D} was significantly associated with gestational diabetes in all participants and participants of European ancestry ($beta_{all}$=0.59, se_{all}=0.063, p_{all}=$1.2 \cdot 10^{-20}$; $beta_{eur}$=0.67, se_{eur}=0.075, p_{eur}=$2.1 \cdot 10^{-19}$) and with PCOS in all participants and participants of European ancestry ($beta_{all}$=0.22, se_{all}=0.043, p_{all}=$1.9 \cdot 10^{-7}$; $beta_{eur}$=0.21, se_{eur}=0.048, p_{eur}=$1.3 \cdot 10^{-5}$) but not with PCOS in participants of African ancestry. These results support a potential genetic basis for the well-known associations between T2D and both gestational diabetes and PCOS[16,17]. In addition, PRS_{T2D} was associated with gestational hypertension in all participants ($beta_{all}$=0.18, se_{all}=0.064, p_{all}=0.0041), breast cancer in all participants ($beta_{all}$=-0.073, se_{all}=0.021, p_{all}=0.00066), and postpartum depression in all participants ($beta_{all}$=0.18, se_{all}=0.07, p_{all}=0.01).

Table 4
Significant (FDR P-value < 0.05) associations between cardiometabolic PRS and female health conditions identified in the PMBB and eMERGE meta-analysis

PRS	Association	Ancestry	Beta	SE	OR	95% CI	P-value
BMI	Breast cancer	All	-0.071	0.0225	0.93	0.891–0.973	0.00159
	Endometrial cancer	All	0.244	0.0458	1.28	1.17–1.4	$9.4 \cdot 10^{-8}$
		EUR	0.278	0.0825	1.32	1.12–1.55	0.000754
	Gestational diabetes	All	0.23	0.0508	1.26	1.14–1.39	$6 \cdot 10^{-6}$
		EUR	0.288	0.0613	1.33	1.18–1.5	$2.59 \cdot 10^{-6}$
	PCOS	All	0.272	0.0387	1.31	1.22–1.42	$2.37 \cdot 10^{-12}$
		EUR	0.281	0.0442	1.32	1.21–1.45	$2.09 \cdot 10^{-11}$
CAD	Breast cancer	All	-0.0718	0.0147	0.931	0.904–0.958	$9.96 \cdot 10^{-7}$
		EUR	-0.0708	0.0156	0.932	0.904–0.961	$5.32 \cdot 10^{-6}$
DBP	Excessive fetal growth	AFR	-0.416	0.13	0.66	0.511–0.851	0.00136
	Gestational hypertension	AFR	0.273	0.0768	1.31	1.13–1.53	0.000372
	Preeclampsia	AFR	0.221	0.0771	1.25	1.07–1.45	0.00421
PP	Gestational hypertension	All	0.218	0.0443	1.24	1.14–1.36	$8.51 \cdot 10^{-7}$
		EUR	0.248	0.0549	1.28	1.15–1.43	$6.6 \cdot 10^{-6}$
	Preeclampsia	All	0.196	0.0581	1.22	1.09–1.36	0.000762
		EUR	0.193	0.0582	1.21	1.08–1.36	0.000914
SBP	Gestational hypertension	All	0.332	0.0825	1.39	1.19–1.64	$5.61 \cdot 10^{-5}$
		EUR	0.31	0.065	1.36	1.2–1.55	$1.77 \cdot 10^{-6}$
		AFR	0.416	0.099	1.52	1.24–1.84	$2.77 \cdot 10^{-5}$
	Preeclampsia	All	0.275	0.0799	1.32	1.13–1.54	0.000567
		EUR	0.234	0.0695	1.26	1.1–1.45	0.000767
		AFR	0.368	0.103	1.44	1.18–1.77	0.00038
T2D	Breast cancer	All	-0.0726	0.0213	0.93	0.892–0.97	0.000657
	Gestational diabetes	All	0.587	0.063	1.8	1.59–2.03	$1.19 \cdot 10^{-20}$
		EUR	0.675	0.0749	1.96	1.7–2.28	$2.13 \cdot 10^{-19}$
	Gestational hypertension	All	0.184	0.0642	1.2	1.06–1.36	0.00414

**BMI = Body Mass Index; CAD = Coronary artery disease, T2D = Type 2 Diabetes; PCOS = Polycystic ovarian syndrome; EUR = European ancestry; P = PMBB; E = eMERGE; Beta coefficients are per SD PRS. Results were adjusted for birth year and the first five PCs.

PRS	Association	Ancestry	Beta	SE	OR	95% CI	P-value
	PCOS	All	0.223	0.0428	1.25	1.15–1.36	$1.93 \cdot 10^{-7}$
		EUR	0.21	0.048	1.23	1.12–1.36	$1.26 \cdot 10^{-5}$
	Postpartum depression	All	0.18	0.07	1.2	1.04–1.37	0.0101

**BMI = Body Mass Index; CAD = Coronary artery disease, T2D = Type 2 Diabetes; PCOS = Polycystic ovarian syndrome; EUR = European ancestry; P = PMBB; E = eMERGE; Beta coefficients are per SD PRS. Results were adjusted for birth year and the first five PCs.

The three PRSs for blood pressure traits (PRS_{DBP}, PRS_{SBP}, and PRS_{PP}) showed varying associations with gestational hypertension and preeclampsia. In the meta-analysis of both cohorts, PRS_{DBP} was significantly associated with gestational hypertension and preeclampsia in participants of African ancestry (gestational hypertension: $beta_{afr}$=0.27, se_{afr}=0.077, p_{afr}=0.00037; preeclampsia: $beta_{afr}$=0.22, se_{afr}=0.077, p_{afr}=0.0042). PRS_{DBP} was also significantly associated with excessive fetal growth in participants of African ancestry ($beta_{afr}$=-0.42, se_{afr}=0.13, p_{afr}=0.0014). PRS_{PP} was significantly associated with gestational hypertension and preeclampsia in all participants and participants of European ancestry (gestational hypertension: $beta_{all}$=0.22, se_{all}=0.044, $p_{all}=8.5 \cdot 10^{-7}$; $beta_{eur}$=0.25, se_{eur}=0.055, $p_{eur}=6.6 \cdot 10^{-6}$; preeclampsia: $beta_{all}$=0.2, se_{all}=0.058, p_{all}=0.00076; $beta_{eur}$=0.19, se_{eur}=0.058, p_{eur}=0.00091). PRS_{SBP} was associated with gestational hypertension and preeclampsia in all participants, participants of European ancestry, and participants of African ancestry in the meta-analysis (gestational hypertension: $beta_{all}$=0.33, se_{all}=0.083, $p_{all}=5.6 \cdot 10^{-5}$; $beta_{eur}$=0.31, se_{eur}=0.065, $p_{eur}=1.8 \cdot 10^{-6}$; $beta_{afr}$=0.42, se_{afr}=0.099, $p_{afr}=2.8 \cdot 10^{-5}$; preeclampsia: $beta_{all}$=0.28, se_{all}=0.08, p_{all}=0.00057; $beta_{eur}$=0.23, se_{eur}=0.07, p_{eur}=0.00077; $beta_{afr}$=0.37, se_{afr}=0.1, p_{afr}=0.00038).

For each significant association, we explored the case prevalence of female-specific health condition per the associated PRS quintile in each cohort (Fig. 2B **and Supplementary Figs. 7, 8, 9, 10, 11, 12, 13, 14, and 15**). The trends matched the results of the association analyses across ancestry groups and in both cohorts. For the positive associations, such as that between PCOS and PRS_{BMI}, the number of cases increased as the PRS percentile increased, whereas the opposite was true for the negative association between breast cancer and PRS_{CAD} (Fig. 2B).

Association between PRS_{CAD} and breast cancer

To validate the inverse association between PRS_{CAD} and breast cancer, we examined the genetic correlation between CAD and breast cancer using the UKBB Genetic Correlation Browser (Fig. 2A). The genetic correlation between I9_CHD (Major coronary heart disease event) and C50 (Diagnoses-main ICD10: C50 Malignant neoplasm of breast) was significantly negative (R_g=-0.24, p = 0.0325). We also compared the effects of SNPs in the CAD GWAS and a large publicly available breast cancer GWAS (**Supplementary Table 2**). Among the 220 SNPs that had a significant effect in the CAD GWAS after pruning, 125 had an opposite effect on breast cancer, and the beta estimates of the SNPs were negatively correlated between the two GWASs (Fig. 2C).

Additionally, to evaluate the risk of ascertainment biases and reporting of comorbidities in the EHR, we assessed the association between PRS_{CAD} and breast cancer in females who were not diagnosed with CAD (n = 61,201). In these individuals, we observed a slightly less significant association between breast cancer and CAD ($beta_{all}$=-0.0484, p_{all}=0.0045; $beta_{eur}$=-0.0457, p_{eur}=0.0115) than in the full dataset. The overlap between the participants diagnosed with CAD and those diagnosed with breast cancer in our cohorts suggests that females with a diagnosis of CAD were less likely to be diagnosed with breast cancer (Fig. 2D).

Mendelian randomization

We performed one-sample MR in the PMBB and eMERGE cohorts for the 29 significant associations and combined them in a meta-analysis (Fig. 3A). The majority of associations remained significant ($p < 0.05$), and the direction of the estimated beta coefficients aligned with the effect directions in the PRS association analysis. The MR results supported causal relationships between many of the cardiometabolic phenotypes and female health conditions, with the most significant associations being between T2D and gestational diabetes (All: $p = 1.5 \cdot 10^{-11}$, beta = 1.52, se = 0.23; EUR: $p = 2.9 \cdot 10^{-10}$, beta = 1.76, se = 0.28) and between BMI and PCOS (All: $p = 1.1 \cdot 10^{-8}$, beta = 0.0021, se = 0.00037; EUR: $p = 6.1 \cdot 10^{-8}$, beta = 0.0021, se = 0.0038).

We also performed two-sample MR to leverage the power of larger GWAS of female-specific health conditions (Fig. 3B). Most of the associations remained significant when using the IVW method but became less significant when using the MR Egger and weighted median methods. Some associations were significant by all three methods, however, particularly those between CAD and breast cancer (IVW: beta=-0.0587, se = 0.021, p = 0.00463; MR Egger: beta=-0.102, se = 0.049, p = 0.04; Weighted median: beta=-0.0642, se = 0.028, p = 0.021) and between T2D and gestational diabetes (IVW: beta = 0.613, se = 0.044, $p = 1.73 \cdot 10^{-43}$; MR Egger: beta = 0.656, se = 0.11, $p = 8.8 \cdot 10^{-10}$; Weighted median: beta = 0.635, se = 0.066, $p = 8.12 \cdot 10^{-22}$).

The role of population stratification

Population stratification can confound the results of PRS association and MR analyses. Therefore, we tested the PRS associations with PCs (**Supplementary Table 3**). The high R^2 values for most cardiometabolic PRSs indicated that much of the variance in the PRSs could be explained by the PCs, although the R^2 values were smaller when European and African ancestry groups were considered separately. The PCs explained more of the PRS variance for participants of African ancestry than for participants of European ancestry, which suggests that there was more population stratification in the former subpopulation than in the latter, and that there may be confounding factors influencing some results. To overcome these biases, we accounted for PCs in both the PRS association analyses and the one-sample MR analyses. Additionally, we performed one-sample MR without adjusting for covariates (PCs and birth year) and compared the results with those of the adjusted MR analysis (**Supplementary Table 4**).

Chronology analyses

We next asked whether genetic risk of cardiometabolic phenotypes affected the time at which female-specific health conditions developed. We found that participants in the younger age groups with high-risk cardiometabolic PRSs tended to have more health conditions than those with low-risk cardiometabolic PRSs (Fig. 4 **and Supplementary Figs. 16, 17, 18, 19, and 20**). In particular, pregnancy-related conditions were relatively more prevalent in participants < 25 years of age with high PRSs than low PRSs for all cardiometabolic PRSs. This difference was less pronounced in participants 25–40 years of age. Although many of the cardiometabolic PRSs did not affect most pregnancy-related conditions, high-risk PRSs that had an effect on pregnancy-related conditions tended to have larger effects in participants < 25 years of age than in participants 25–40 years of age. In addition, PRSs for high blood pressure tended to have greater effects than other PRSs on gestational hypertension and preeclampsia in the younger age groups (< 25 and 25–40 years). We observed similar trends for PRS_{T2D} and gestational diabetes and for PRS_{BMI} and PCOS. We also observed high prevalence of endometriosis and uterine fibrosis in participants with high-risk PRSs, particularly those 25–40 years of age.

Discussion

We observed genetic correlations between cardiometabolic traits and a variety of obstetric and gynecological disorders. Additionally, we showed that genome-wide PRSs for cardiometabolic traits have potential translational value for individuals with female-specific health conditions. The genetic correlations between cardiometabolic phenotypes and female-specific health conditions suggest a high degree of shared genetic etiology among all the phenotypes tested in this study. We investigated the effects of this shared genetic etiology by generating PRSs for cardiometabolic phenotypes and testing their associations with various female-specific health conditions. The PRSs were generally predictive of their corresponding cardiometabolic phenotypes as well as several female-specific health conditions. Prior research based mainly on non-genetic factors provided evidence for some of these associations, such those between BMI and PCOS and between T2D and gestational diabetes[15,16]. Epidemiological studies have also provided evidence of a positive correlation between obesity and endometrial cancer[40]. Our results suggest that certain relationships between cardiometabolic phenotypes and female-specific health conditions have a genetic basis. Furthermore, our MR results suggest that some of these relationships are causal.

PRS_{CAD} was inversely associated with breast cancer in individuals of European ancestry. CAD and breast cancer share many common risk factors, such as smoking and poor diet[41]; however, our results indicated that a high genetic risk of CAD was protective against breast cancer, as participants with high PRS_{CAD} had lower incidence of breast cancer than participants with low PRS_{CAD}. When we considered only participants with high PRS_{CAD} but no CAD diagnosis, we still saw a moderately reduced risk of breast cancer. Several factors might contribute to these results. First, the patterns might reflect ascertainment biases and competing risks. Individuals with high risk of CAD likely live shorter lives than individuals with low risk of CAD and are therefore less likely to be diagnosed with breast cancer during their lifetime. Second, treatments for CAD might protect against breast cancer. Third, genetic mechanisms that predispose individuals to CAD might have protective effects against breast cancer. Our results, together with those of another recent study that showed a protective effect of PRS_{CAD} on breast cancer, suggest that more research should be undertaken to uncover the mechanisms behind this association[42].

The differences among the three-blood pressure related PRSs suggests that analysis of the genetic risk of all three blood pressure traits might improve the understanding of hypertension during pregnancy and elucidate interesting differences due to genetic burden in individuals of different ancestries. PRS_{DBP} was significantly associated with gestational hypertension and preeclampsia in only participants of African ancestry, while only PRS_{SBP} was significantly associated with gestational hypertension and preeclampsia in all participants and participants of European and African ancestry. Prior studies identified variability in predictive performance when different blood pressure measurements were used to predict hypertension and other diseases. Our results warrant further studies to replicate this association in other large multi-ancestry datasets to reach a more definite conclusion on the role blood pressure measurements play in predicting hypertensive diseases during pregnancy[43,44].

High genetic risks of most of the cardiometabolic phenotypes tended to increase the risk of certain health conditions at a young age, even in the absence of an overall association. For example, PRS_{BMI} was not associated with gestational hypertension and ectopic pregnancy overall; however, in the youngest age group (< 25 years), these complications were more prevalent in participants with high PRS_{BMI} than in participants with low PRS_{BMI}. Similarly, there was no overall association between PRS_{SBP} and PCOS or endometriosis, but participants who developed PCOS before 25 years of age or endometriosis at 25–40 years of age were more likely to have high PRS_{SBP} than low PRS_{SBP}. The difference in the prevalence of these conditions between participants with high-risk PRS and participants with low-risk PRS became smaller with increasing age. Patients with high genetic risk of cardiometabolic phenotypes may therefore have an elevated risk of early development of female-specific health conditions. These findings may help to prioritize patients for early screening.

Our study establishes a link between the genetic risks of cardiometabolic phenotypes and several diseases that are unique to females of European and African ancestry; however, we estimated the genetic risk using PRS alone, which is based on common variants and does not include the effects of other rare variants and copy-number variations. In addition, we did not consider clinical or environmental factors, such as education level and socioeconomic status. We found that population stratification was potentially present in our cohorts, so we accounted for PCs in our analyses accordingly. Our focus on PRSs reflects the limitations of methods for multi-modal risk prediction. Current interest in building integrative risk models suggests that this gap can be closed in the near future[25,45]. Furthermore, the weaker performance of PRSs in individuals of African ancestry contributed to a lack of power to identify associations specific to those individuals and suggests a need for future studies to include more racially and ethnically diverse cohorts.

The small sample size for some conditions may have limited the power of our one-sample MR analysis. Although our two-sample MR used GWASs with larger cohorts, these cohorts were primarily of European ancestry and therefore lacked the power to reveal causality in individuals of non-European ancestry. Furthermore, the causal effects identified in our MR analyses might be biased because of horizontal pleiotropy, in which genetic variants associated with cardiometabolic phenotypes affect other traits that in turn influence female-specific health conditions. We included methods such as MR Egger that account for pleiotropy and other confounding factors but may not have captured all their effects.

EHRs are particularly advantageous for investigating disease trajectories and progression. Our analysis provided a visualization of the burden of early diagnosis of female-specific health conditions in individuals with high genetic risk of cardiometabolic phenotypes. However, these results might have been affected by ascertainment biases of EHRs and inclusion of confounding factors.

This study illustrates the influence of genetic risk of cardiometabolic phenotypes on female-specific health conditions and highlight differences in the genetic predisposition among individuals of European and African ancestries. Our findings suggest that cardiometabolic PRSs may have clinical utility as non-modifiable risk factors for screening and early diagnosis of a variety of obstetric and gynecological conditions. To improve the power of cardiometabolic PRSs to predict risks of female-specific health conditions, future studies should incorporate PRSs with other genetic and non-genetic risk factors and study their effects on larger and more diverse multi-ancestry populations.

Declarations

Acknowledgements: We acknowledge the Penn Medicine BioBank (PMBB) for providing data and thank the patient-participants of Penn Medicine who consented to participate in this research program. We would also like to thank the Penn Medicine BioBank team and Regeneron Genetics Center for providing genetic variant data for analysis. The PMBB is approved under IRB protocol# 813913 and supported by Perelman School of Medicine at University of Pennsylvania.

Sources of Funding: The eMERGE Network was initiated and funded by NHGRI through the following grants:
Phase IV: U01HG011172 (Cincinnati Children's Hospital Medical Center); U01HG011175 (Children's Hospital of Philadelphia); U01HG008680 (Columbia University); U01HG011176 (Icahn School of Medicine at Mount Sinai); U01HG008685 (Mass General Brigham); U01HG006379 (Mayo Clinic); U01HG011169 (Northwestern University); U01HG011167 (University of Alabama at Birmingham); U01HG008657 (University of Washington Medical Center, Seattle); U01HG011181 (Vanderbilt University Medical Center); U01HG011166 (Vanderbilt University Medical Center serving as the Coordinating Center)
Phase III: U01HG8657 (Kaiser Permanente Washington/University of Washington Medical Center); U01HG8685 (Brigham and Women's Hospital); U01HG8672 (Vanderbilt University Medical Center); U01HG8666 (Cincinnati Children's Hospital Medical Center); U01HG6379 (Mayo Clinic); U01HG8679 (Geisinger Clinic); U01HG8680 (Columbia University Health Sciences); U01HG8684 (Children's Hospital of Philadelphia); U01HG8673 (Northwestern University); U01HG8701

(Vanderbilt University Medical Center serving as the Coordinating Center); U01HG8676 (Partners Healthcare/Broad Institute); and U01HG8664 (Baylor College of Medicine).

Disclosures: None

Author Contributions: B.X. performed the analysis. S.S.V designed the study. B.X. and S.S.V wrote the manuscript. T.D, A.L helped with data extraction and analyses. H.H, L.L, N.E, W.Q.W, Y.L helped with data generation. B.X, D.V.E, T.D, K.G. B.L, C.W. G.P.J, D.K, M.D.R, S.S.V critically reviewed and approved the manuscript. Regeneron Genetics Center genotyped the Penn Medicine BioBank dataset.

References

1. Chang C, Zhang K, Veluchamy A, et al. A Genome-Wide Association Study Provides New Evidence That CACNA1C Gene is Associated With Diabetic Cataract. Invest Ophthalmol Vis Sci. 2016;57(4):2246–2250.
2. Patel SR. Shared genetic risk factors for obstructive sleep apnea and obesity. J Appl Physiol (1985). 2005;99(4):1600–1606.
3. Zhuang QS, Zheng H, Gu XD, Shen L, Ji HF. Detecting the genetic link between Alzheimer's disease and obesity using bioinformatics analysis of GWAS data. Oncotarget. 2017;8(34):55915–55919.
4. Rashkin SR, Graff RE, Kachuri L, et al. Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. Nat Commun. 2020;11(1):4423. Published 2020 Sep 4.
5. Blüher M. Obesity: global epidemiology and pathogenesis. Nat Rev Endocrinol. 2019;15(5):288–298.
6. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract*. 2019;157:107843.
7. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015 [published correction appears in Lancet. 2017 Jan 7;389(10064):e1]. *Lancet*. 2016;388(10053):1459–1544.
8. Gerdts E, Regitz-Zagrosek V. Sex differences in cardiometabolic disorders. Nat Med. 2019;25(11):1657–1666.
9. Buddeke J, Bots ML, van Dis I, et al. Comorbidity in patients with cardiovascular disease in primary care: a cohort study with routine healthcare data. Br J Gen Pract. 2019;69(683):e398-e406.
10. Kendir C, van den Akker M, Vos R, Metsemakers J. Cardiovascular disease patients have increased risk for comorbidity: A cross-sectional study in the Netherlands. Eur J Gen Pract. 2018;24(1):45–50.
11. Kulie T, Slattengren A, Redmer J, Counts H, Eglash A, Schrager S. Obesity and women's health: an evidence-based review. J Am Board Fam Med. 2011;24(1):75–85.
12. Fox R, Kitt J, Leeson P, Aye CYL, Lewandowski AJ. Preeclampsia: Risk Factors, Diagnosis, Management, and the Cardiovascular Impact on the Offspring. J Clin Med. 2019;8(10):1625.
13. Garovic VD, August P. Preeclampsia and the future risk of hypertension: the pregnant evidence. Curr Hypertens Rep. 2013;15(2):114–121.
14. Tooher J, Thornton C, Makris A, Ogle R, Korda A, Hennessy A. All Hypertensive Disorders of Pregnancy Increase the Risk of Future Cardiovascular Disease. Hypertension. 2017 Oct;70(4):798–803.
15. Gambineri A, Pelusi C, Vicennati V, Pagotto U, Pasquali R. Obesity and the polycystic ovary syndrome. Int J Obes Relat Metab Disord. 2002;26(7):883–896.

16. Kwak SH, Jang HC, Park KS. Finding genetic risk factors of gestational diabetes. Genomics Inform. 2012;10(4):239–243.
17. Gambineri A, Patton L, Altieri P, et al. Polycystic ovary syndrome is a risk factor for type 2 diabetes: results from a long-term prospective study. Diabetes. 2012;61(9):2369–2374. doi:10.2337/db11-1360
18. Mu F, Rich-Edwards J, Rimm EB, Spiegelman D, Missmer SA. Endometriosis and Risk of Coronary Heart Disease. Circ Cardiovasc Qual Outcomes. 2016;9(3):257–264.
19. Dawson A, Fernandez ML, Anglesio M, Yong PJ, Carey MS. Endometriosis and endometriosis-associated cancers: new insights into the molecular mechanisms of ovarian cancer development. Ecancermedicalscience. 2018;12:803.
20. Paul A, Paul S. The breast cancer susceptibility genes (BRCA) in breast and ovarian cancers. Front Biosci (Landmark Ed). 2014;19:605–618.
21. Nilsson E, Salonen Ros H, Cnattingius S, Lichtenstein P. The importance of genetic and environmental effects for pre-eclampsia and gestational hypertension: a family study. BJOG. 2004;111(3):200–206.
22. Vink JM, Sadrzadeh S, Lambalk CB, Boomsma DI. Heritability of polycystic ovary syndrome in a Dutch twin-family study. J Clin Endocrinol Metab. 2006;91(6):2100–2104. doi:10.1210/jc.2005-1494
23. Vujkovic M, Keaton JM, Lynch JA, et al. 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. 2020;52(7):680–691.
24. Justice AE, Winkler TW, Feitosa MF, et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. Nat Commun. 2017;8:14977.
25. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Med. 2020;12(1):44. Published 2020 May 18.
26. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219–1224.
27. Steinthorsdottir V, McGinnis R, Williams NO, et al. Genetic predisposition to hypertension is associated with preeclampsia in European and Central Asian women. Nat Commun. 2020;11(1):5976.
28. Verma A, Ritchie MD. Current Scope and Challenges in Phenome-Wide Association Studies. Curr Epidemiol Rep. 2017;4(4):321–329.
29. Fritsche LG, Gruber SB, Wu Z, et al. Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from The Michigan Genomics Initiative. Am J Hum Genet. 2018;102(6):1048–1061.
30. Shen X, Howard DM, Adams MJ, et al. A phenome-wide association and Mendelian Randomisation study of polygenic risk for depression in UK Biobank. Nat Commun. 2020;11(1):2301.
31. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.
32. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47(3):291–295.
33. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467(7311):52–58.
34. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51(4):584–591.
35. Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun. 2019;10(1):3328.
36. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software. 2010;36(3):1–48.

37. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019;10(1):1776.
38. Pendergrass SA, Dudek SM, Crawford DC, Ritchie MD. Visually integrating and exploring high throughput Phenome-Wide Association Study (PheWAS) results using PheWAS-View. BioData Min. 2012;5(1):5.
39. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408.
40. Onstad MA, Schmandt RE, Lu KH. Addressing the Role of Obesity in Endometrial Cancer Risk, Prevention, and Treatment. J Clin Oncol. 2016;34(35):4225–4230.
41. Mehta LS, Watson KE, Barac A, et al. Cardiovascular Disease and Breast Cancer: Where These Entities Intersect: A Scientific Statement From the American Heart Association [published correction appears in Circulation. 2019 Aug 27;140(9):e543]. *Circulation*. 2018;137(8):e30-e66.
42. Clark SL, Parham M, Shadyab AH, et al. Broad Clinical Manifestations of Polygenic Risk for Coronary Artery Disease in the Women's Health Initiative. medRxiv 2021.06.15.21258993
43. Kanegae H, Oikawa T, Okawara Y, Hoshide S, Kario K. Which blood pressure measurement, systolic or diastolic, better predicts future hypertension in normotensive young adults?. J Clin Hypertens (Greenwich). 2017;19(6):603–610.
44. Sesso HD, Stampfer MJ, Rosner B, et al. Systolic and diastolic blood pressure, pulse pressure, and mean arterial pressure as predictors of cardiovascular disease risk in Men. Hypertension. 2000;36(5):801–807.
45. Kachuri L, Graff RE, Smith-Byrne K, et al. Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. Nat Commun. 2020;11(1):6084.

Figures

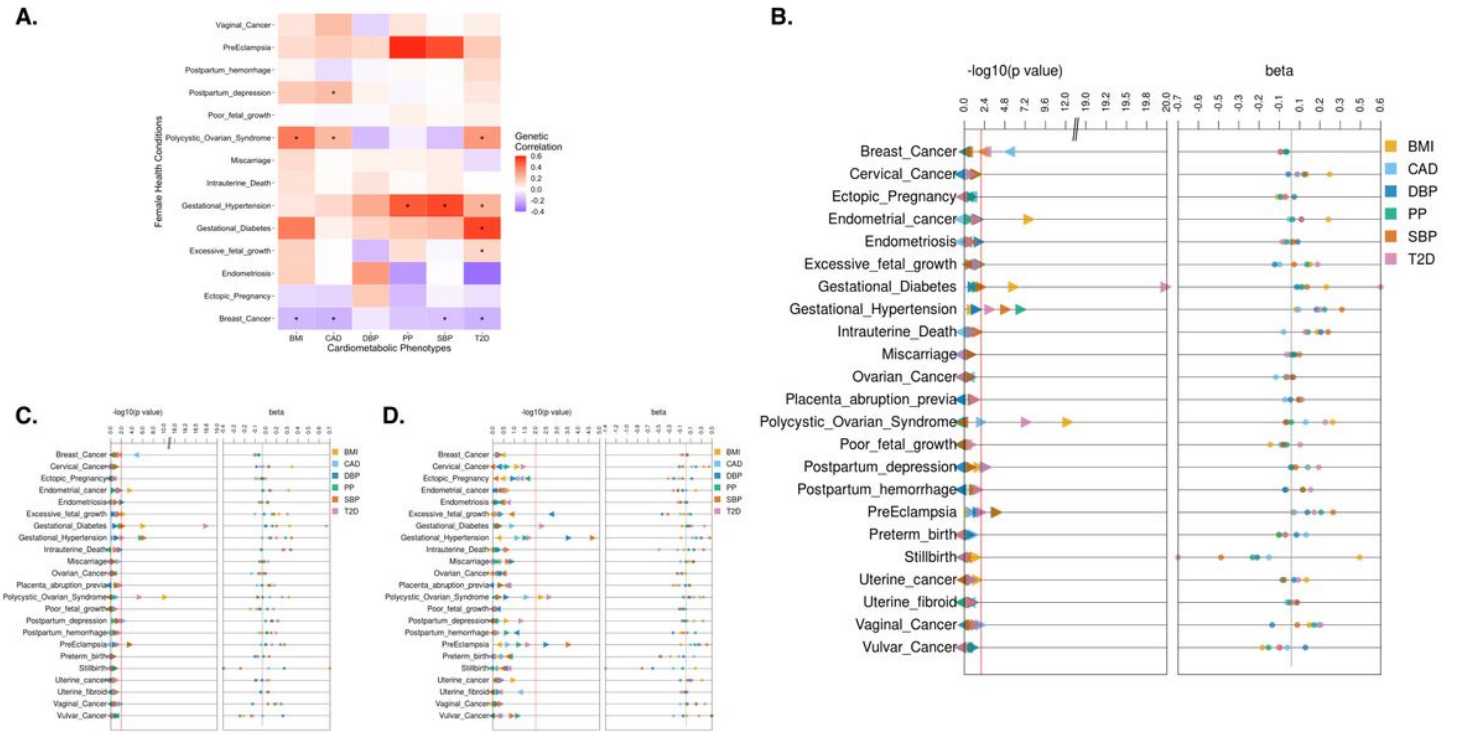

Figure 1

Genetic correlation and the influence of shared genetic burden of cardiometabolic traits and health conditions unique to females. Panel A shows a heatmap of genetic correlations from cross-trait LDSC analyses: Blue represents negative correlation and red represents positive correlation. The asterisk in each box indicates statistical significance of results. Panels B, C, and D show results of PRS-based meta-analyses between cardiometabolic PRSs and case/control status of female-specific diseases overall (Panel B) and in participants of European (Panel C) and African (Panel D) ancestry, respectively. The first panel in these plots corresponds to raw, unadjusted p-values, and the second panel shows beta estimates. The color of each point refers to the PRS for cardiometabolic traits. The red line corresponds to the p = 0.01 threshold.

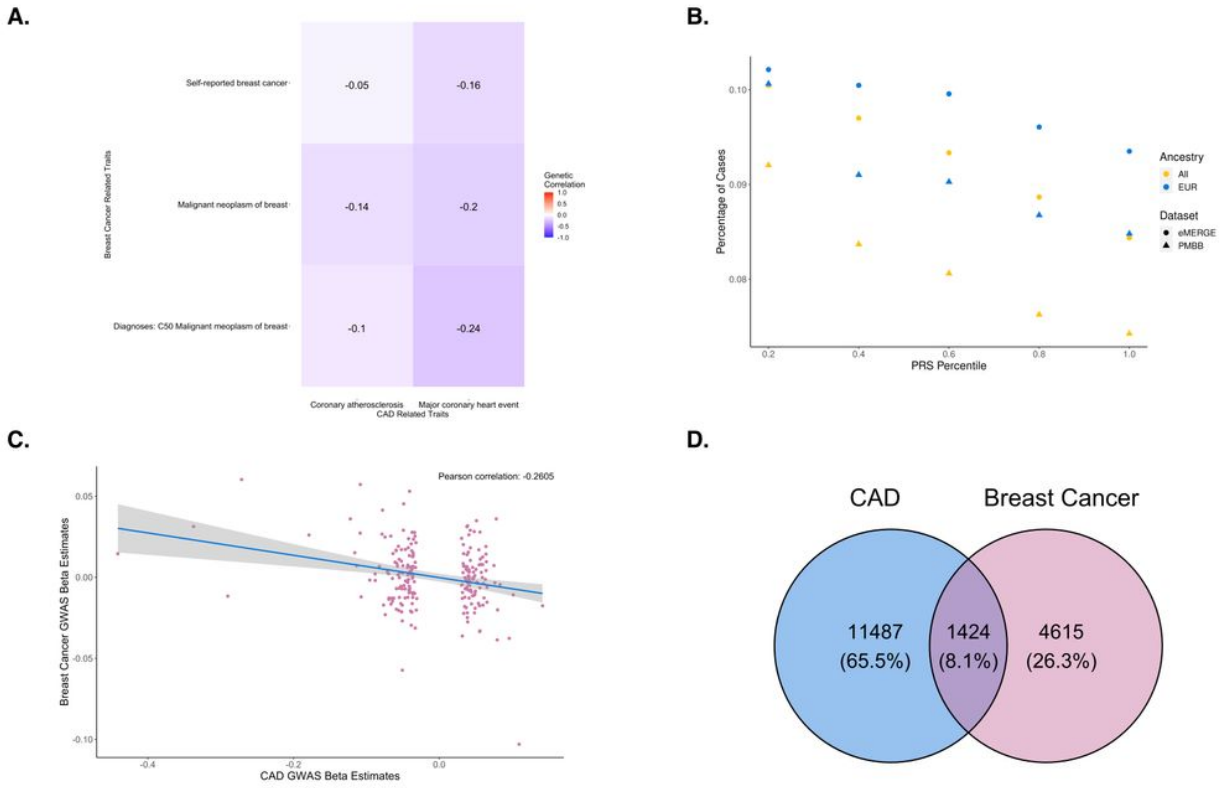

Figure 2

Inverse relationship between coronary artery disease (CAD) and breast cancer. Panel A shows a heatmap of negative genetic correlations between CAD and breast cancer from the UKBB Genetic Correlation Browser dataset. The gradient of color shows positive (red) to negative (blue) correlations, and the text in each box gives the genetic correlation coefficient. Panel B shows the distribution of breast cancer per each PRS_{CAD} quintile. The x-axis represents each PRS quintile, and the y-axis represents the disease prevalence. The color of each point refers to the ancestry group (All, European, and African), and the shape indicates the target dataset (eMERGE [circle] or PMBB [triangle]). Panel C shows the effect estimates of the genome-wide significant GWAS SNPs in the CAD GWAS and the corresponding effect estimates of the same SNPs in the breast cancer GWAS. The blue line is the line of best fit from linear regression, and the shaded area around the line is the 95% confidence interval. Panel D is a Venn diagram representing the overlap of CAD and breast cancer cases in the PMBB and eMERGE datasets.

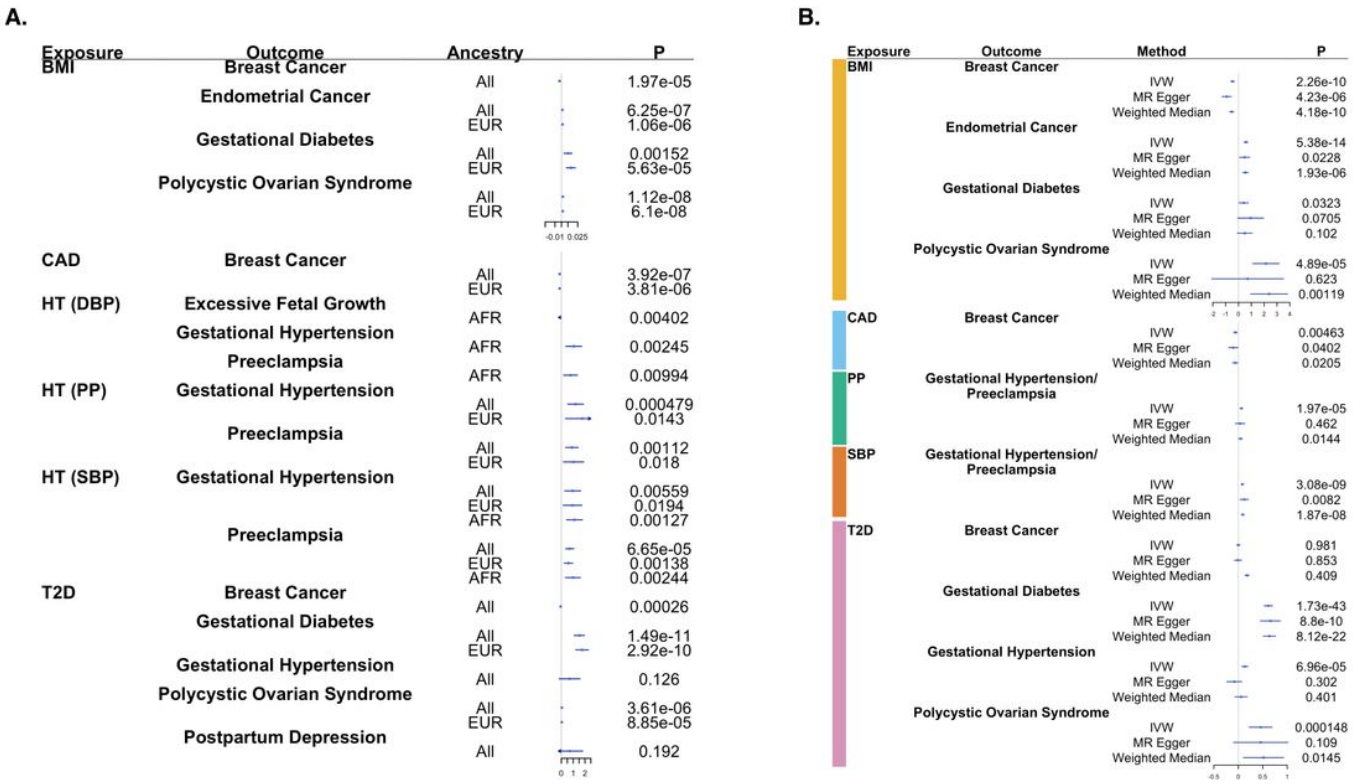

Figure 3

Mendelian randomization (MR) results for 30 significant PRS-based associations. Forest plots show associations between female health conditions as outcomes and cardiometabolic phenotypes as exposures in a one-sample (A) and two-sample (B) MR analyses. Genetic instruments are the cardiometabolic PRS in the one-sample MR and genome-wide significant SNPs from GWAS in the two-sample MR. In panel A, each point refers to beta outcome/SD exposure for PRS_{BMI} and log(OR) outcome for all other exposures for each test performed as separated by ancestry. In panel B, each point refers to beta outcome/SD exposure for BMI and beta outcome/SD log(OR) exposure for all other variables across all methods used in sensitivity analyses for two-sample MR. P-values are reported in the last column in both panels.

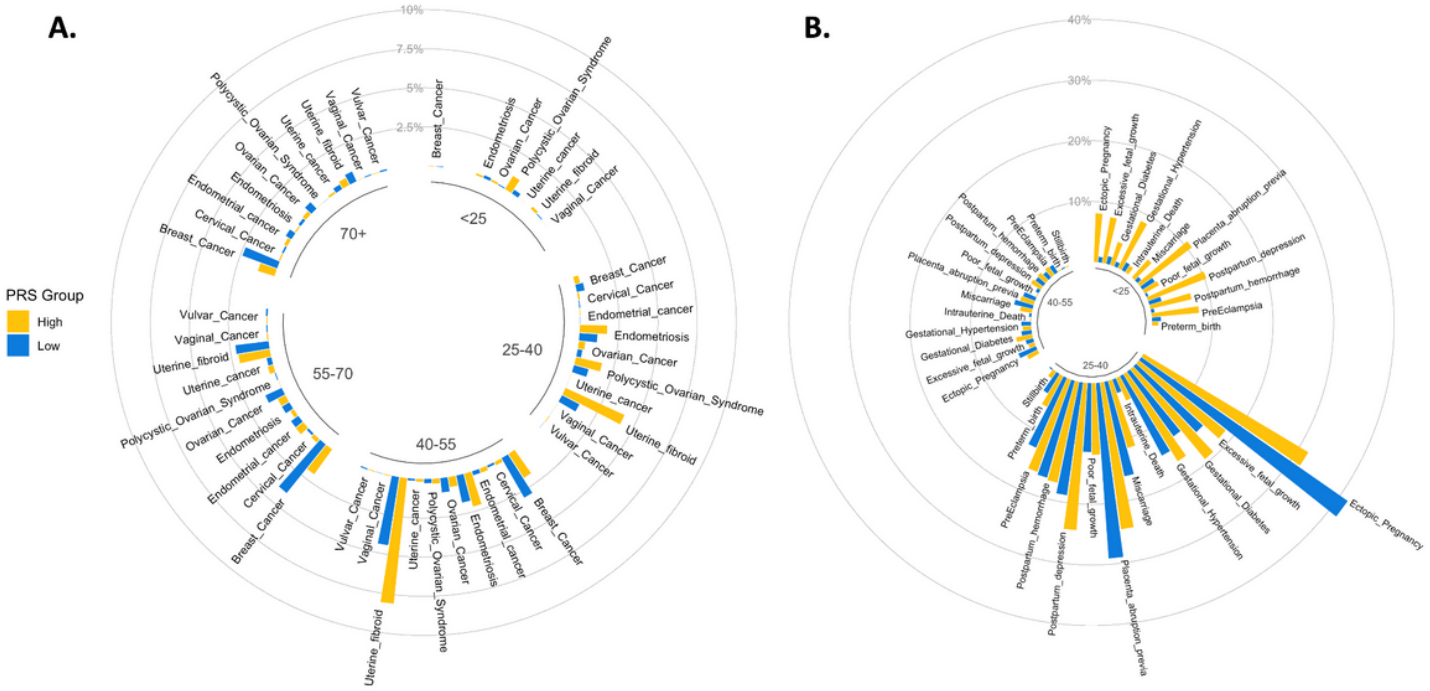

Figure 4

Chronology analyses of events from the electronic health records. A circular plot shows disease prevalence among participans with high PRS_{BMI} (in yellow) and low PRS_{BMI} (in blue). General female health conditions are shown in panel A, and pregnancy and childbirth-related phenotypes are shown in panel B. The circular plots are divided into five age categories (<25, 25–40, 40–55, 55–70, and >70) for general female health conditions and three age categories (<25, 25–40, and 40–55) for pregnancy-related phenotypes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- NatureCommunicationsSupplementaryTables.xlsx
- NatureCommunicationsSupplementaryFigures.pdf
- rs.pdf