Journal of the American Heart Association

ORIGINAL RESEARCH

Inference of Causal Relationships Between Genetic Risk Factors for Cardiometabolic Phenotypes and Female-Specific Health Conditions

Brenda Xiao, BS; Digna R. Velez Edwards, PhD, MS; Anastasia Lucas, BS; Theodore Drivas, MD, PhD; Kathryn Gray, MD, PhD; Brendan Keating, PhD; Chunhua Weng, PhD; Gail P. Jarvik, MD, PhD; Hakon Hakonarson, MD, PhD; Leah Kottyan, PhD; Noemie Elhadad, PhD; Wei-Qi Wei, MD, PhD; Yuan Luo, PhD; Dokyoon Kim, PhD; Marylyn Ritchie, PhD; Shefali Setia Verma, PhD; for the Regeneron Genetics Center*

BACKGROUND: Cardiometabolic diseases are highly comorbid, but their relationship with female-specific or overwhelmingly female-predominant health conditions (breast cancer, endometriosis, pregnancy complications) is understudied. This study aimed to estimate the cross-trait genetic overlap and influence of genetic burden of cardiometabolic traits on health conditions unique to women.

METHODS AND RESULTS: Using electronic health record data from 71 008 ancestrally diverse women, we examined relationships between 23 obstetrical/gynecological conditions and 4 cardiometabolic phenotypes (body mass index, coronary artery disease, type 2 diabetes, and hypertension) by performing 4 analyses: (1) cross-trait genetic correlation analyses to compare genetic architecture, (2) polygenic risk score–based association tests to characterize shared genetic effects on disease risk, (3) Mendelian randomization for significant associations to assess cross-trait causal relationships, and (4) chronology analyses to visualize the timeline of events unique to groups of women with high and low genetic burden for cardiometabolic traits and highlight the disease prevalence in risk groups by age. We observed 27 significant associations between cardiometabolic polygenic scores and obstetrical/gynecological conditions (body mass index and endometrial cancer, body mass index and polycystic ovarian syndrome, type 2 diabetes and gestational diabetes, type 2 diabetes and polycystic ovarian syndrome). Mendelian randomization analysis provided additional evidence of independent causal effects. We also identified an inverse association between coronary artery disease and breast cancer. High cardiometabolic polygenic scores were associated with early development of polycystic ovarian syndrome and gestational hypertension.

CONCLUSIONS: We conclude that polygenic susceptibility to cardiometabolic traits is associated with elevated risk of certain female-specific health conditions.

Key Words: cardiometabolic diseases ■ female health conditions ■ genomic burden ■ Mendelian randomization ■ polygenic risk scores

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.[1–3] Many studies have shown that the pathophysiology of cardiometabolic diseases affects men and women differently.[4] Although

Correspondence to: Shefali Setia Verma, PhD, D202 Richards Building, 3700 Hamilton Walk, University of Pennsylvania, Philadelphia, PA 19104. Email: shefali.setiaverma@pennmedicine.upenn.edu

Preprint posted on MedRxiv February 04, 2022. doi: https://doi.org/10.1101/2022.02.02.22269844.

*A complete list of the Regeneron Genetics Center members can be found in the Appendix at the end of the article.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.121.026561

For Sources of Funding and Disclosures, see page 15.

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

CLINICAL PERSPECTIVE

What Is New?

- In this study of >70 000 women from multiple electronic health record–linked biobanks, we evaluated polygenic scores' ability to measure genetic predispositions for a variety of clinically important traits and disorders unique to women.
- Using multifactorial approaches for evaluating shared genetic burden and causal relationships, we demonstrate the potential for an inverse causal relationship between coronary artery disease and breast cancer.
- We have produced a chronological map of female health including disease prevalence based on shared genetic burden of cardiometabolic traits from adolescence to old age.

What Are the Clinical Implications?

- Women with high genetic burden of cardiometabolic conditions may be predisposed to certain diseases such as gestational hypertension, gestational diabetes, and polycystic ovarian syndrome.

Nonstandard Abbreviations and Acronyms

eMERGE	Electronic Medical Records and Genomics
FDR	false discovery rate
IVW	inverse variance weighted
LD	linkage disequilibrium
MR	Mendelian randomization
PC	principal component
PMBB	Penn Medicine BioBank

cardiometabolic diseases have many sequelae that affect both sexes, including depression, anxiety, chronic obstructive pulmonary disease, and cancer,[5,6] they disproportionately affect women because of their links to female-specific health conditions such as preeclampsia, gestational diabetes, stillbirth, and pregnancy loss.[7,8] Multiple lines of evidence link female health conditions to cardiometabolic diseases, where the incidence of cardiometabolic conditions, such as obesity and T2D, leads to associated morbidities. Individuals who develop preeclampsia during pregnancy are more likely to develop cardiovascular diseases and hypertension after pregnancy.[9,10] Obesity is closely linked to polycystic ovarian syndrome (PCOS), and individuals with PCOS are at high risk of developing T2D.[11–13] Individuals with endometriosis are at high risk of developing various cancers and cardiovascular diseases such as myocardial infarction and ischemic heart disease.[14,15] 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 (GWASs) have exposed common genetic causes among diseases such as T2D, obesity, sleep apnea, hypertension, Alzheimer disease, and many types of cancer.[16–19] There is evidence of impacts on cardiometabolic phenotypes for >1000 different loci, but the effect size of any single variant is generally minimal. To estimate an individual's overall risk of disease, researchers have used GWAS results to calculate polygenic scores (PGSs), which sum the effects of common single-nucleotide polymorphisms (SNPs) on a given phenotype[20] and can be included in models predicting the risks of many cardiometabolic phenotypes and comorbidities.[21,22]

To investigate the effects of genetic risk of cardiometabolic phenotypes on female health conditions, we measured PGSs for cardiometabolic phenotypes and then determined their correlations with female-specific health conditions documented in electronic health records (EHRs). Prior studies successfully used PGSs as genetic risk factors linking multiple adverse phenotypes.[23,24] PGS-based association tests have the advantage that they are based on unvarying risk factors (ie, inherited genetic burden), and they make fewer assumptions than association tests based on individual genetic factors. We hypothesized that cardiometabolic phenotypes and female-specific health conditions share common genetic factors so that genetic risks for adverse cardiometabolic phenotypes should associate with 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 PMBB (Penn Medicine BioBank) and the eMERGE (Electronic Medical Records and Genomics) 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 associations between the genetic risks of adverse cardiometabolic phenotypes and female-specific health conditions. In addition, we evaluated potential causal relationships between 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 scores to understand age-specific prevalence in high-risk and low-risk individuals.

METHODS

The authors declare that all supporting data for these analyses will be available within the article through its online supplementary files. The authors have individual-level access to genotype data and medical records from the PMBB and eMERGE network data sets. The eMERGE data set is available for download on dbGAP (phs001584.v2. p2). The PMBB can be accessed upon reasonable request to the corresponding author. The PMBB study was approved by University of Pennsylvania's institutional review board, and the subjects gave informed consent.

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. The PMBB is approved under institutional review board protocol number 813913 and is supported by the Perelman School of Medicine at the University of Pennsylvania. All subjects in the PMBB gave informed consent. The PMBB links patient genotype data with detailed EHR information. Currently, the PMBB contains imputed genotype data for ≈45 000 samples. The PMBB cohort is diverse, with >25% of its participants being of African ancestry. PMBB genotype data were imputed to the The TransOmics for Precesion Medicine (TOPMed) Reference panel using the Michigan Imputation server. We included 21 837 participants from PMBB in this study who self-reported their sex as women (Table 1). The stratified analyses in this study included only participants of European or African ancestry, whereas those of Asian or Hispanic ancestry were excluded because of sample size limitations. We defined the genetically informed ancestry for PMBB participants by using the kernel density estimation on the principal components.

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. All participants included in eMERGE gave informed consent, and the study was approved by each institute's respective institutional review board. 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 self-reported 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. The ancestry of eMERGE participants was determined based on the intersection of self-reported race and principal component-based k-means clustering.[25]

Statistical Analysis

GWASs of Cardiometabolic Phenotypes

To determine genetic correlations and calculate PGSs, we collected quantitative data on the effects of genetic variants on cardiometabolic phenotypes from several large GWASs. Given the diverse nature of our study population, we obtained the largest publicly available multiancestry GWAS summary statistics for 4 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.

Genome-Wide Associations of Female-Specific Health Conditions

We included 23 female-specific health conditions to evaluate in this study, including breast cancer, which can affect both sexes but is far more prevalent in women than men. Large multiancestry GWASs are not available for most of these conditions. Therefore, we used PLINK (version 1.90) to conduct GWASs of female-specific health conditions documented in the PMBB and eMERGE data sets.[30] We filtered the genetic variants in the PMBB and eMERGE imputed data sets to include only variants with imputation quality R^2>0.3 and minor allele frequency >0.01. We then conducted separate GWASs for European and African ancestry individuals in the 2 cohorts and combined results from the 2 cohorts using the meta-analysis command in PLINK.

Genetic Correlation Calculation

We used the GWASs obtained and generated above to calculate pairwise genetic correlations between cardiometabolic phenotypes and female-specific health conditions using linkage disequilibrium score regression.[31] Linkage disequilibrium score accounts for linkage disequilibrium (LD) among SNPs by using an external reference panel that should match the ancestry distribution

Table 1. Sample Sizes in the PMBB and eMERGE Data Sets 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	20209	12344	6744	39403	32880	5555
CAD	21837 (3002)	13515 (1956)	7039 (955)	49171 (9597)	38918 (7953)	9067 (1515)
DBP	21612	13343	6994	NA	NA	NA
Hypertension	21837 (10278)	13515 (5640)	7039 (4286)	49171 (27685)	38918 (21883)	9067 (5265)
T2D	21837 (4388)	13515 (1969)	7039 (2221)	49171 (12403)	38918 (8999)	9067 (3090)
Female health phenotypes						
Breast cancer	21837 (1621)	13515 (1146)	7039 (415)	49171 (4148)	38918 (3639)	9067 (443)
Cervical cancer	21837 (105	13515 (53)	7039 (50)	49171 (332)	38918 (263)	9067 (60)
Ectopic pregnancy	2808 (1779)	1201 (827)	1319 (757)	3078 (823)	1975 (570)	641 (172)
Endometrial cancer	21837 (286)	13515 (183)	7039 (90)	49171 (771)	38918 (680)	9067 (83)
Endometriosis	21837 (701)	13515 (320)	7039 (340)	49171 (2314)	38918 (1891)	9067 (354)
Excessive fetal growth	693 (85)	293 (37)	329 (42)	2433 (627)	1618 (422)	437 (162)
Gestational diabetes	2655 (523)	1111 (193)	1262 (248)	3174 (762)	2005 (445)	719 (251)
Gestational hypertension	2666 (631)	1118 (256)	1275 (324)	3135 (703)	1980 (411)	711 (253)
Intrauterine death	685 (57)	289 (23)	324 (27)	2156 (64)	1438 (44)	358 (19)
Miscarriage	2934 (389)	1273 (199)	1360 (151)	3568 (823)	2323 (577)	740 (200)
Ovarian cancer	21837 (305)	13515 (191)	7039 (89)	49171 (1589)	38918 (1359)	9067 (197)
Placenta abruption/ previa	1192 (396)	482 (157)	586 (195)	2674 (997)	1697 (672)	583 (262)
Polycystic ovarian syndrome	21837 (736)	13515 (387)	7039 (272)	49171 (1006)	38918 (750)	9067 (209)
Poor fetal growth	792 (202)	325 (78)	388 (107)	2209 (167)	1478 (115)	360 (27)
Postpartum depression	924 (384)	374 (136)	466 (226)	2417 (555)	1579 (349)	457 (177)
Postpartum hemorrhage	846 (283)	350 (108)	408 (146)	2320 (401)	1544 (261)	408 (109)
Preeclampsia	2631 (452)	1100 (149)	1264 (272)	3132 (702)	1974 (406)	713 (249)
Preterm birth	687 (66)	284 (21)	332 (40)	2144 (57)	1433 (35)	350 (16)
Stillbirth	649 (17)	270 (3)	311 (12)	2123 (8)	1417 (7)	347 (0)
Uterine cancer	21837 (113)	13515 (66)	7039 (43)	49171 (418)	38918 (348)	9067 (64)
Uterine fibroid	21837 (1570)	13515 (516)	7039 (984)	49171 (5711)	38918 (4103)	9067 (1455)
Vaginal cancer	21837 (23)	13515 (13)	7039 (9)	49171 (109)	38918 (89)	9067 (18)
Vulvar cancer	21837 (41)	13515 (25)	7039 (14)	49171 (120)	38918 (97)	9067 (21)

BMI indicates body mass index; CAD, coronary artery disease; DBP, diastolic blood pressure; eMERGE, Electronic Medical Records and Genomics; PMBB, Penn Medicine BioBank; and T2D, type 2 diabetes.

of the corresponding GWASs. We generated European and African ancestry LD reference panels using the HapMap3 SNPs (≈1 M common variants) from all members included in the respective 1000 Genomes population.[32] We then used these LD reference panels to calculate the genetic correlation separately in European and African ancestry participants. We then combined correlation results across ancestries through a meta-analysis using the metafor R package under the restricted maximum likelihood model.

Polygenic Scores

PGSs calculated using a GWAS of 1 particular ancestry group tend to perform poorly when applied to individuals from a different ancestry group.[20,33,34] To accurately calculate PGSs for our diverse cohorts, we calculated PGSs 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 April 24, 2020), a method that performs Polygenic Prediction via Bayseian

Table 2. GWAS Data Sets Used to Calculate Genetic Correlations and Polygenic Scores and the Number of SNPs Used From Each GWAS to Calculate the Corresponding PGS

Phenotype	Ancestries included	Source	Sample size (N cases)	PMID	No. SNPs included in PGS
Type 2 diabetes	EUR, AFR, EAS, SAS, HIS	Ref. [26]	1 407 282 (228 499)	32 541 925	PMBB: 1 023 697 eMERGE: 716 330
Body mass index	EUR, AFR, EAS, SAS, HIS	Ref. [27]	241 258	28 443 625	PMBB: 885 143 eMERGE: 614 668
Hypertension (DBP, SBP, PP)	EUR, AFR, EAS, SAS, HIS	Ref. [28]	318 891	30 578 418	PMBB: 1 024 567 eMERGE: 715 471
Coronary artery disease	EUR, AFR, EAS, SAS, HIS	Ref. [29]	547 261 (122 733)	29 212 778	PMBB: 981 480 eMERGE: 681 029

AFR indicates African ancestry; CARDIoGRAMplusC4D, Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) pus The Coronary Artery Disease)C4D) Genetics Consortium; DBP, diastolic blood pressure; EAS, East Asian ancestry; eMERGE, Electronic Medical Records and Genomics; EUR, European ancestry; GIANT, Genetic Investigation of Anthropometric Traits; GWAS, genome-wide association study; HIS, Hispanic ancestry; MVP, Million Veterans Program; PGS, polygenic score; PMBB, Penn Medicine BioBank; PP, pulse pressure; SAS, South Asian ancestry; SBP, systolic blood pressure; SNPs, single-nucleotide polymorphisms; and UKBB, UK BioBank.

Regression and Continous Shrinkage Priors.[35] PRS-CS requires a reference panel that matches the ancestry distribution of the target data set. We generated multiple reference panels for analyses: European-only reference panel from 1000 Genomes European ancestry population, African-only reference panel from 1000 Genomes African ancestry population, and a multiancestry LD reference panel using the HapMap SNPs from the entire 1000 Genomes populations (2504 individuals). We identified LD patterns within the 1000 Genomes population by using PLINK (version 1.90) to determine LD blocks and calculate the LD between the SNPs in each block. For PRS-CS, the global shrinkage parameter φ was fixed to 0.01, and default values were selected for all other parameters. PGSs were then calculated using the weights with PLINK. Only the SNPs in the target data set, summary statistics, and LD reference panel were included in the PGSs. The numbers of SNPs used for each PGS calculation are listed in Table 2. The scores were then normalized (mean of 0 and standard deviation of 1) for each analysis separately (stratified by ancestry and overall).

To evaluate the power of the PGSs, we tested their performance on association of the corresponding primary phenotypes in the summary statistics. We could not obtain quantitative measurements of blood pressure for all participants in the eMERGE cohort, and PP and SBP measurements were not curated in the PMBB cohort. Therefore, we evaluated the performance of the PGSs for blood pressure traits (SBP, DBP, PP) in the eMERGE cohort based on hypertension case–control phenotypes, and we evaluated the performance of the PGS 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 PGS performance by the area under the receiver operating curves 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. We also evaluated the value of including PGS in our models by using the likelihood ratio test to compare the null (covariate only) model with the full (PGS and covariates) model using the lmtest package in R. The regression models used birth year and the first 5 principal components (PCs) as covariates. PCs for PMBB and eMERGE were determined from projection onto the 1000 Genomes population. We tested PGS performance overall as well as for subgroups of European or African ancestry. In addition, we constructed European and African LD reference panels using individuals from the respective ancestry groups from 1000 Genomes and computed PGSs using these ancestry-specific LD panels. We then compared the performance of the PGSs with the PGSs generated using the multiancestry LD reference panel in the PMBB cohort.

Phenotype Data

Cases and controls for each phenotype were defined using *International Classification of Diseases* (*ICD-9 and ICD-10*) diagnosis codes. Participants were coded as cases of a given phenotype if their records contained at least 1 of the corresponding *ICD-9 or ICD-10* codes. For pregnancy-related phenotypes, participants were only considered controls if their records had at least 1 pregnancy-related *ICD-9 or ICD-10* code and no *ICD-9 or ICD-10* 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-9 or ICD-10* code. The complete list of *ICD-9 and ICD-10* codes used to include or exclude participants as cases and controls can be found in Table S1. Using these definitions, we determined the sample size for each phenotype in the eMERGE and PMBB cohorts (Table 1).

PGS Association Analysis

We tested the association between each cardiometabolic PGS and female-specific health condition by fitting separate logistic regression models adjusted by birth year and the first 5 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 Benjamini-Hochberg false discovery rate (FDR) significance threshold of 0.05, adjusting the *P* values with the number of hypotheses tested (6 cardiometabolic PGSs×23 female-specific health conditions=138 hypotheses). The logistic regressions were performed using the glm function in R. The results from the PMBB and eMERGE cohorts were meta-analyzed 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.[37] We then created prevalence plots for each significant association. The participants were divided into quintiles based on the PGS 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 1-sample MR for 27 significant associations from the PGS analyses using the ivreg function in the ivpack R package and tested which relationships were significant after Benjamini-Hochberg FDR correction (adjusting for 27 hypotheses). Cardiometabolic PGSs we calculated before were used as genetic instruments, because effect sizes for each SNP were adjusted according to the significance of the association through PRS-CS. We also included birth year and the first 5 PCs as covariates. The results were then combined in a meta-analysis using the metafor R package under the restricted maximum-likelihood estimator model. Because the small sample size for some of the conditions could limit the power of the analysis, we also performed 2-sample MR for the same associations using the inverse variance weighted (IVW) method in the twoSampleMR package in R.[38] MR sensitivity analyses were conducted using the weighted median and MR Egger methods in the same package, and tests for horizontal pleiotropy were also performed using MR Egger. The genetic instruments for the cardiometabolic phenotypes in the 2-sample MR were the genome-wide significant SNPs ($P<5\times10^{-8}$) in the respective cardiometabolic GWAS used to calculate the PGSs. The SNPs were pruned according to LD patterns among the 1000 Genomes HapMap SNPs (r^2=0.1, kb=250), and the remaining representative SNPs were included in the analysis. Matching genetic instruments for the female-specific health conditions were obtained from publicly available GWASs through the twoSampleMR package (Table S2). Bidirectional MR was also performed for female-specific health conditions that had at least 1 genome-wide significant SNP in the GWAS after pruning (r^2=0.1, kb=250). Because these GWASs for female-specific health conditions were conducted in only European ancestry populations, we performed 2-sample MR for only associations that were significant in all participants or participants of European ancestry. Two-sample MR results were then tested for significance after multiple hypothesis correction using an FDR threshold of 0.05 (adjusting for 13 unique significantly associated female-specific health conditions). Genetic instrument strength for each analysis was calculated using the mean *F* statistic (β^2/σ^2) across all SNPs from the cardiometabolic GWASs included in the MR.

Chronology Analyses

We divided the participants into high-risk and low-risk groups according to each cardiometabolic PGS. High PGS was defined as the top quintile (>80th percentile), and low PGS was defined as the bottom quintile (<20th percentile). We considered age at the first occurrence of each female-specific health condition from the *ICD-9 or ICD-10* records. For pregnancy-related conditions, the participants were split into 3 age groups: <25, 25 to 39, and 40 to 55 years. We excluded participants who were >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 nonpregnancy related conditions, the participants were split into 5 age groups: <25, 25 to 39, 40 to 54, 55 to 69, and ≥70 years. We then examined the combined case prevalence of each health condition within the high and low PGS groups in each cohort across all age groups.

RESULTS

Genetic Correlations Among Cardiometabolic Phenotypes and Female-Specific Health Conditions

We calculated heritability estimates for 13 female-specific health conditions and proceeded with the genetic correlation analysis for these conditions. We found 10 correlations between the cardiometabolic phenotypes and 6 unique female-specific health conditions that were at least nominally significant ($P<0.05$) after meta-analysis of ancestry-specific results. (Figure 1A). Two correlations were FDR significant: SBP positively with gestational hypertension (R_g=0.292, P=0.0013 [all *P* values reported in the text are raw *P* values]) and T2D with PCOS (R_g=0.158, $P=5.1\times10^{-6}$). BMI was nominally positively correlated with postpartum depression (R_g=0.102, P=0.049), CAD with PCOS (R_g=0.148, P=0.0047), PP with gestational hypertension (R_g=0.244, P=0.0058) and preeclampsia (R_g=0.165, P=0.043), SBP

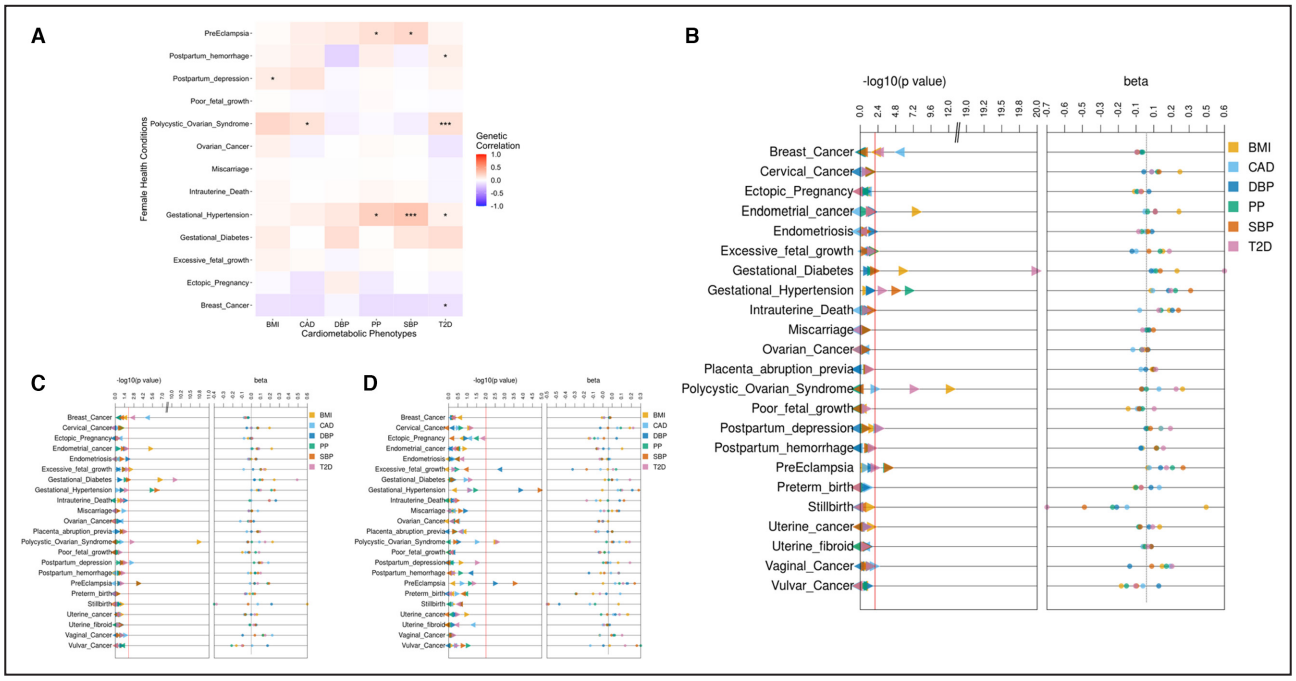


Figure 1. Genetic correlation and the influence of shared genetic burden of cardiometabolic traits and health conditions unique to women.

A, Heatmap of genetic correlations between cardiometabolic phenotypes and female-specific health conditions using multiancestry cardiometabolic genome-wide association studies (GWASs) and European and African ancestry meta-analyzed GWASs for female-specific health conditions. Blue represents negative correlation and red represents positive correlation. Triple asterisks in a box indicate false discovery rate significance, and a single asterisk in a box indicates nominal significance (P<0.05). Genetic correlation was unable to be calculated for grayed-out boxes. **B** through **D**, Results of polygenic score (PGS)-based meta-analyses between cardiometabolic PGSs and case–control status of female-specific health conditions overall (**B**) and in participants of European (**C**) and African (**D**) ancestry. The first panel in these plots corresponds to raw, unadjusted P values, and the second panel shows β estimates. The color of each point refers to the PGS for cardiometabolic traits. The red line corresponds to the P=0.01 threshold. BMI indicates body mass index, CAD, coronary artery disease; DBP, diastolic blood pressure; PP; pulse pressure; SBP, systolic blood pressure; and T2D, type 2 diabetes.

with preeclampsia (R_g=0.202, P=0.0089), and T2D with excessive fetal growth (R_g=0.0331, P=0.03) and gestational hypertension (R_g=0.0711, P=0.014). T2D was also negatively correlated with breast cancer (R_g=−0.126, P=0.048). Some other correlations were significant in only 1 ancestry group (Figure S1). For example, CAD was significantly negatively correlated with breast cancer (R_g=−0.241, P=0.0013) and T2D was significantly positively correlated with gestational diabetes (R_g=0.256, P=4.5×10^{-9}) in the European ancestry-specific GWASs, and T2D was nominally positively correlated with postpartum depression (R_g=0.0773, P=0.035) and postpartum hemorrhage (R_g=0.13, P=0.0067) in the African ancestry-specific GWASs.

PGS Performance for Predicting Primary Phenotypes

We calculated a PGS for each cardiometabolic phenotype and confirmed the distributions of the raw and normalized scores (Figures S2 through S7). The full model for all PGSs 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 PGS-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 PGSs in European and African ancestry individuals and compared these differences between these 2 groups. The PGSs significantly 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 PGSs were more accurate in individuals with European ancestry than in individuals with African ancestry, but they improved models in both groups. PGS performance when using different ancestry LD reference panels was similar across phenotypes and ancestry groups (Table S3). Because of relatively more consistent performance and the multiancestry nature of our GWASs and target data sets, we chose to use the PGSs calculated using the multiancestry reference panel for further analyses.

Table 3. Effect Estimates for Tests of Association of PGS With Primary Phenotype in the eMERGE and PMBB Data Sets

Phenotype	Ancestry	N total	N cases	R^2 or AUC of full model	β	SE	*P* value	R^2 or AUC of null model	R^2 or AUC of PGS only model	Likelihood test *P* value
eMERGE										
BMI	All	39338		0.08482	2.378	0.05044	0	0.0334	0.0593	0
	EUR	32880		0.07584	1.91	0.04046	0	0.0132	0.0622	0
	AFR	5555		0.07118	1.585	0.1358	4.02E-31	0.0485	0.0239	3.62E-31
CAD	All	49171	9597	0.784	0.3091	0.01318	1.62E-121	0.7745	0.5474	1.35E-123
	EUR	38918	7953	0.7732	0.3044	0.01404	2.7E-104	0.7624	0.5629	1.07E-106
	AFR	9067	1515	0.8228	0.2161	0.03193	1.3E-11	0.8197	0.5462	1.06E-11
DBP, hypertension	All	49171	27685	0.8109	0.1387	0.01347	7.17E-25	0.8099	0.529	5.99E-25
	EUR	38918	21883	0.7994	0.1346	0.01294	2.48E-25	0.798	0.5355	1.95E-25
	AFR	9067	5265	0.8619	0.1308	0.03028	1.56E-05	0.8612	0.5068	1.5E-05
PP, hypertension	All	49171	27685	0.8115	0.1644	0.01344	2.15E-34	0.8099	0.526	1.5E-34
	EUR	38918	21883	0.8	0.1454	0.01235	5.75E-32	0.798	0.5276	3.88E-32
	AFR	9067	5265	0.8613	0.059	0.02987	0.0484	0.8612	0.5044	0.0483
SBP, hypertension	All	49171	27685	0.8131	0.2977	0.01665	1.86E-71	0.8099	0.5352	3.97E-72
	EUR	38918	21883	0.8023	0.253	0.01427	2.7E-70	0.798	0.5446	3.08E-71
	AFR	9067	5265	0.8622	0.1817	0.03525	2.55E-07	0.8612	0.5031	2.36E-07
T2D	All	49171	12403	0.7193	0.6334	0.01914	3.02E-240	0.691	0.6163	1.68E-259
	EUR	38918	8999	0.6975	0.4924	0.01515	7.23E-232	0.655	0.6145	1.18E-244
	AFR	9067	3090	0.7808	0.3786	0.03372	3.07E-29	0.7724	0.5483	9.11E-30
PMBB										
BMI	All	20209		0.1542	2.658	0.08244	2.05E-222	0.1108	0.1417	1.66E-222
	EUR	12344		0.07977	1.9	0.06055	6.14E-208	0.006422	0.07283	4.43E-208
	AFR	6744		0.03556	1.696	0.122	2.46E-43	0.008037	0.2969	2.18E-43
CAD	All	21837	3002	0.8235	0.3571	0.02354	5.73E-52	0.8148	0.5619	8.5E-53
	EUR	13515	1956	0.8303	0.3958	0.02794	1.46E-45	0.8185	0.5929	3.13E-47
	AFR	7039	955	0.799	0.1887	0.03832	8.51E-07	0.7955	0.5389	7.89E-07
DBP	All	21612		0.0425	1.036	0.06661	3.05E-54	0.03183	0.02905	2.91E-54
	EUR	13343		0.01555	0.7934	0.06362	1.71E-35	0.004144	0.0124	1.62E-35
	AFR	6994		0.04545	0.7797	0.09169	2.23E-17	0.03571	0.0164	2.13E-17
PP, hypertension	All	21837	10278	0.8125	0.1942	0.02328	7.18E-17	0.8112	0.602	5.22E-17
	EUR	13515	5640	0.7797	0.1474	0.02008	2.11E-13	0.7778	0.5272	1.73E-13
	AFR	7039	4286	0.839	0.1205	0.03334	0.0003	0.8383	0.5391	0.00029
SBP, hypertension	All	21837	10278	0.8144	0.327	0.02523	2.08E-38	0.8112	0.6097	4.24E-39
	EUR	13515	5640	0.7817	0.2205	0.0204	5.72E-25	0.7778	0.5397	2.84E-25
	AFR	7039	4286	0.8416	0.2763	0.03502	3.01E-15	0.8383	0.5649	1.65E-15
T2D	All	21837	4388	0.754	0.8136	0.03523	5.49E-118	0.7287	0.6713	8.59E-130
	EUR	13515	1969	0.7133	0.5425	0.02701	9.4E-90	0.6634	0.6323	5.9E-95
	AFR	7039	2221	0.7366	0.5051	0.03602	1.13E-44	0.7168	0.5919	3.89E-46

The β coefficients and SEs are per SD PGS in the full model. Covariates included birth year and the first 5 principal components. AFR indicates African ancestry; AUC, area under the receiver operating curve; BMI, body mass index; CAD, coronary artery disease; DBP, diastolic blood pressure; eMERGE, Electronic Medical Records and Genomics; EUR, European ancestry; PGS, polygenic score; PMBB, Penn Medicine BioBank; PP, pulse pressure; SBP, systolic blood pressure; and T2D, type 2 diabetes.

Associations Between Cardiometabolic Risks and Female-Specific Health Conditions

We detected numerous associations between cardiometabolic PGSs and female health conditions in the meta-analysis of both cohorts (Figure 1B through 1D). Twenty-seven associations were statistically significant after correction for multiple hypothesis testing (Table 4). In the meta-analysis, for all participants and participants of European ancestry, PGS_{BMI} was significantly positively associated with endometrial cancer (β_{all}=0.24, SE_{all}=0.046, P_{all}=9.4×10^{-8}; β_{eur}=0.21, SE_{eur}=0.047, P_{eur}=1.01×10^{-5}) and gestational diabetes (β_{all}=0.23, SE_{all}=0.051, P_{all}=6×10^{-6}; β_{eur}=0.23, SE_{eur}=0.046, P_{eur}=4.4×10^{-7}). PGS_{BMI} was also inversely associated with breast cancer in all participants (β_{all}=−0.071, SE_{all}=0.022, P_{all}=0.0016) and positively associated with PCOS (β_{all}=0.27, SE_{all}=0.039, P_{all}=2.4×10^{-12}). 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. PGS_{CAD} was negatively associated with breast cancer for all participants and participants of European ancestry (β_{all}=−0.072, SE_{all}=0.015, P_{all}=1×10^{-6}; β_{eur}=−0.071, SE_{all}=0.016, P_{eur}=6.5×10^{-6}). This finding suggests that individuals, particularly those of European ancestry, with high PGS_{CAD}, are relatively less likely to have breast cancer compared with individuals with low PGS_{CAD}. PGS_{T2D} was significantly positively associated with gestational diabetes in all participants and participants of European ancestry (β_{all}=0.59, SE_{all}=0.063, P_{all}=1.2×10^{-20}; β_{eur}=0.46, SE_{eur}=0.071, P_{eur}=9.4×10^{-11}) and with PCOS in all participants (β_{all}=0.22, SE_{all}=0.043, P_{all}=1.9×10^{-7}). These results support a potential genetic basis for the well-known associations between T2D and

Table 4. Significant (False Discovery Rate *P*<0.05) Associations Between Cardiometabolic PGS and Female Health Conditions Identified in the PMBB and eMERGE Meta-Analysis

PGS	Association	Ancestry	β	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	9.4×10^{-8}
		EUR	0.206	0.0465	1.23	1.12–1.35	1.01×10^{-5}
	Gestational diabetes	All	0.23	0.0508	1.26	1.14–1.39	6×10^{-6}
		EUR	0.232	0.0458	1.26	1.15–1.38	4.36×10^{-7}
	PCOS	All	0.272	0.0387	1.31	1.22–1.42	2.37×10^{-12}
		EUR	0.214	0.032	1.24	1.16–1.32	2.13×10^{-11}
CAD	Breast cancer	All	−0.0718	0.0147	0.931	0.904–0.958	9.96×10^{-7}
		EUR	−0.0707	0.0157	0.932	0.904–0.961	6.52×10^{-6}
DBP	Excessive fetal growth	AFR	−0.313	0.0976	0.731	0.604–0.886	0.00135
	Gestational hypertension	AFR	0.204	0.0545	1.23	1.1–1.36	0.000179
PP	Gestational hypertension	All	0.218	0.0443	1.24	1.14–1.36	8.51×10^{-7}
		EUR	0.204	0.0452	1.23	1.12–1.34	6.62×10^{-6}
	Preeclampsia	All	0.196	0.0581	1.22	1.09–1.36	0.000762
		EUR	0.165	0.0486	1.18	1.07–1.3	0.000661
SBP	Gestational hypertension	All	0.332	0.0825	1.39	1.19–1.64	5.61×10^{-5}
		EUR	0.233	0.0482	1.26	1.15–1.39	1.32×10^{-6}
		AFR	0.261	0.0605	1.3	1.15–1.46	1.63×10^{-5}
	Preeclampsia	All	0.275	0.0799	1.32	1.13–1.54	0.000567
		EUR	0.165	0.0486	1.18	1.07–1.3	0.000661
		AFR	0.225	0.0631	1.25	1.11–1.42	0.000357
T2D	Breast cancer	All	−0.0726	0.0213	0.93	0.892–0.97	0.000657
		EUR	−0.0556	0.017	0.946	0.915–0.978	0.00109
	Gestational diabetes	All	0.587	0.063	1.8	1.59–2.03	1.19×10^{-20}
		EUR	0.46	0.0711	1.58	1.38–1.82	9.37×10^{-11}
	Gestational hypertension	All	0.184	0.0642	1.2	1.06–1.36	0.00414
	PCOS	All	0.223	0.0428	1.25	1.15–1.36	1.93×10^{-7}

P values reported are the raw *P* values for all associations that are false discovery rate significant. Results were adjusted for birth year and the first 5 principal components. The β coefficients are per SD PGS. BMI indicates body mass index; CAD, coronary artery disease; DBP, diastolic blood pressure; eMERGE, Electronic Medical Records and Genomics; EUR, European ancestry; OR, odds ratio; PGS, polygenic score; PCOS, polycystic ovarian syndrome; PMBB, Penn Medicine BioBank; SBP, systolic blood pressure; and T2D, type 2 diabetes.

both gestational diabetes and PCOS.[12,13] In addition, PGS_{T2D} was positively associated in with gestational hypertension in all participants (β_{all}=0.18, SE_{all}=0.064, P_{all}=0.0041) and negatively associated with breast cancer in all participants and participants of European ancestry (β_{all}=−0.073, SE_{all}=0.021, P_{all}=0.00066; β_{eur}=−0.056, SE_{eur}=0.017, P_{eur}=0.0011).

The 3 PGSs for blood pressure traits (PGS_{DBP}, PGS_{SBP}, and PGS_{PP}) showed varying associations with gestational hypertension and preeclampsia. In the meta-analysis of both cohorts, PGS_{DBP} was significantly positively associated with gestational hypertension in participants of African ancestry (gestational hypertension: β_{afr}=0.2, SE_{afr}=0.055, P_{afr}=0.00018). PGS_{DBP} was also negatively associated with excessive fetal growth in participants of African ancestry (β_{afr}=−0.31, SE_{afr}=0.098, P_{afr}=0.0014). PGS_{PP} was positively associated with gestational hypertension and preeclampsia in all participants and participants of European ancestry (gestational hypertension: β_{all}=0.22, SE_{all}=0.044, P_{all}=8.5 ×10^{-7}; β_{eur}=0.2, SE_{eur}=0.045, P_{eur}=6.6×10^{-6}; preeclampsia: β_{all}=0.2, SE_{all}=0.058, P_{all}=0.00076; β_{eur}=0.17, SE_{eur}=0.049, P_{eur}=0.00066). PGS_{SBP} was positively associated with gestational hypertension and preeclampsia in all participants, participants of European ancestry, and participants of African ancestry (gestational hypertension: β_{all}=0.33, SE_{all}=0.083, P_{all}=5.6×10^{-5}; β_{eur}=0.23, SE_{eur}=0.048, P_{eur}=1.3×10^{-6}; β_{afr}=0.26, SE_{afr}=0.061, P_{afr}=1.6×10^{-5}; preeclampsia: β_{all}=0.28, SE_{all}=0.08, P_{all}=0.00057; β_{eur}=0.17, SE_{eur}=0.049, P_{eur}=0.00066; β_{afr}=0.23, SE_{afr}=0.063, P_{afr}=0.00036).

For each significant association, we explored the case prevalence of female-specific health condition per the associated PGS quintile in each cohort (Figure 2A and Figures S8 through S14). The trends matched the results

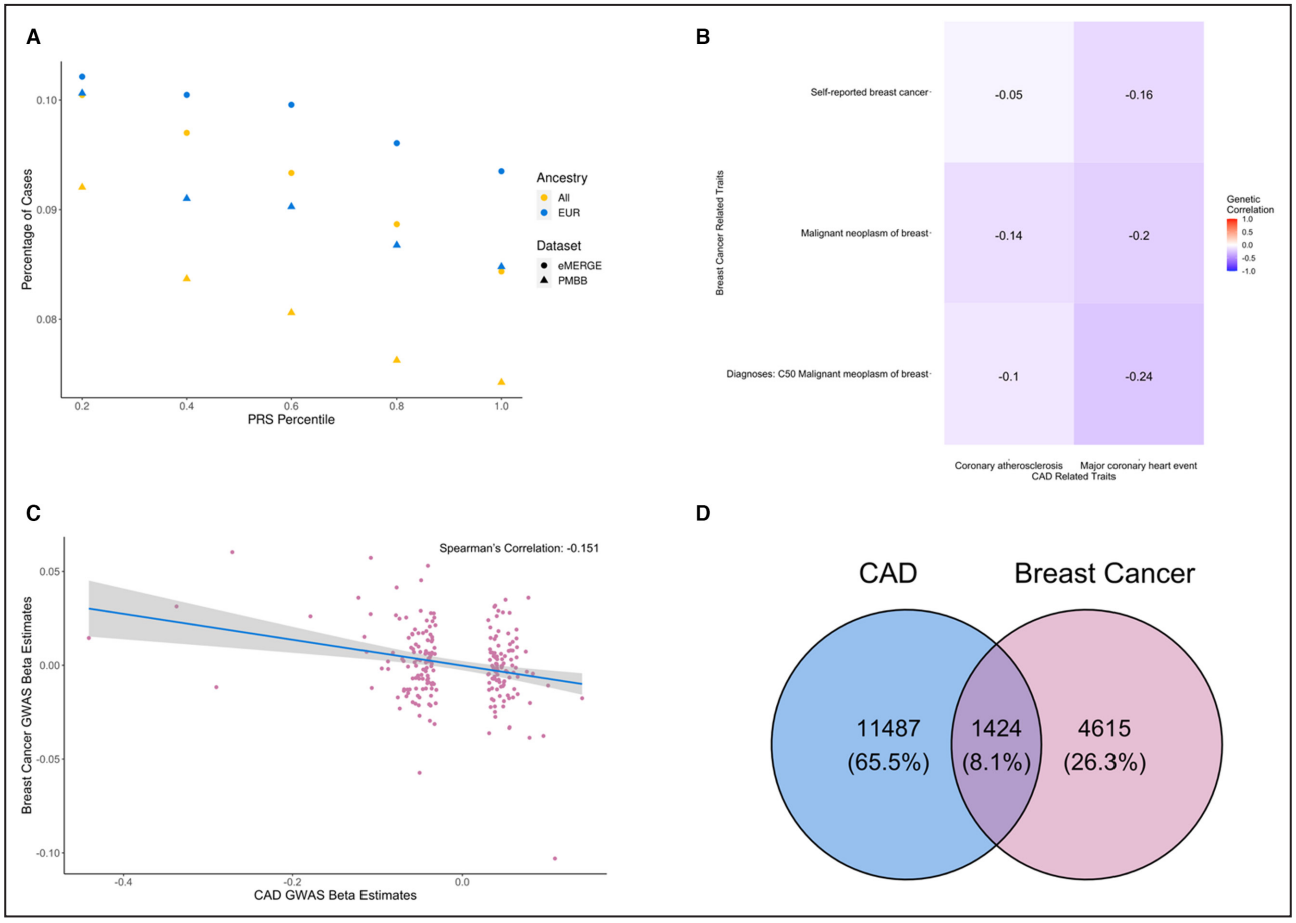


Figure 2. Inverse relationship between coronary artery disease (CAD) and breast cancer.
A, Distribution of breast cancer per each polygenic score $(PGS)_{CAD}$ quintile. The *x* axis represents each PGS quintile, and the *y* axis represents the disease prevalence. The color of each point refers to the ancestry group (All, European [EUR], and African [AFR]), and the shape indicates the target data set (eMERGE [●] or PMBB [▲]). **B**, Heatmap of negative genetic correlations between CAD and breast cancer from the UK BioBank Genetic Correlation Browser data set. The gradient of color shows positive (red) to negative (blue) correlations, and the text in each box gives the genetic correlation coefficient. **C**, Effect estimates of the genome-wide significant genome-wide association studies (GWAS) single nucleotide polymorphisms (SNPs) in the CAD GWASs and the corresponding effect estimates of the same SNPs in the breast cancer GWASs. The blue line is the line of best fit from linear regression, and the shaded area around the line is the 95% CI. **D**, Venn diagram representing the overlap of CAD and breast cancer cases in the PMBB and eMERGE data sets. eMERGE indicates Electronic Medical Records and Genomics; and PMBB, Penn Medicine BioBank.

of the association analyses across ancestry groups and in both cohorts. For the positive associations, such as that between PCOS and PGS_{BMI}, the number of cases increased as the PGS percentile increased, whereas the opposite was true for negative associations such as that between breast cancer and PGS_{CAD} (Figure 2A).

Association Between PGS_{CAD} and Breast Cancer

To validate the inverse association between PGS_{CAD} and breast cancer, we examined the genetic correlation between CAD and breast cancer using the UK BioBank Genetic Correlation browser (Figure 2B). The genetic correlation between I9_CHD (Major coronary heart disease event) and C50 (Diagnoses-main *ICD-10*: C50 Malignant neoplasm of breast) was significantly negatively correlated (R_g=−0.24, P=0.0325). We also compared the effects of SNPs in the CAD GWASs and a large publicly available breast cancer GWAS (Table S2). Among the 220 genome-wide significant SNPs in the CAD GWASs after pruning, 125 had an opposite effect on breast cancer, and the β estimates of the SNPs were negatively correlated between the 2 GWASs (R_g=−0.151, Spearman's correlation) (Figure 2C).

Additionally, to evaluate the risk of ascertainment biases and reporting of comorbidities in the EHR, we assessed the association between PGS_{CAD} and breast cancer in women who were not diagnosed with CAD (n=61 201). We observed a weaker but still negative association between breast cancer and CAD (β_{all}=−0.0481, P_{all}=0.0047; β_{eur}=−0.045, P_{eur}=0.0113) than in the full data set, and this association also remained significant. The overlap between the participants diagnosed with CAD and those diagnosed with breast cancer in our cohorts suggests that women with a diagnosis of CAD were less likely to be diagnosed with breast cancer (Figure 2D).

Mendelian Randomization

We performed 1-sample MR in the PMBB and eMERGE cohorts for the 27 significant associations and combined them in a meta-analysis (Figure 3A). The majority of associations were nominally significant (raw

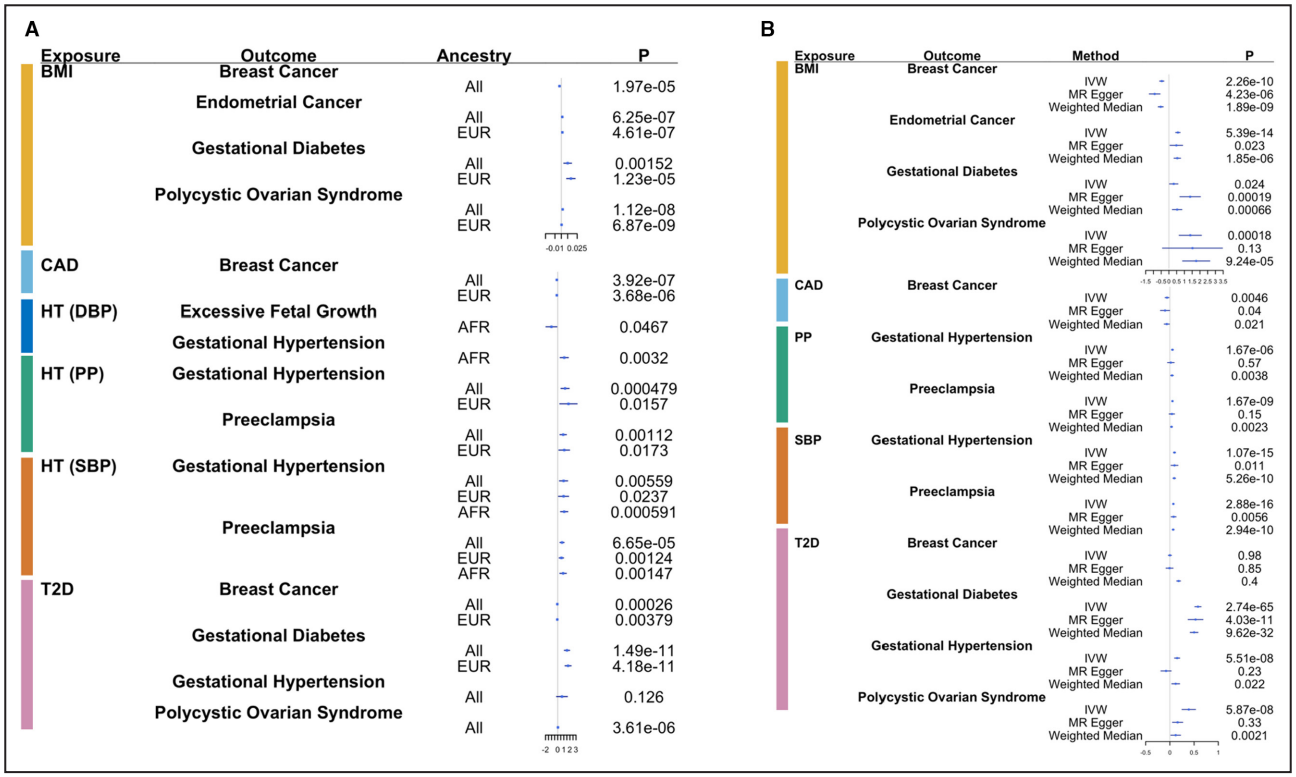


Figure 3. Mendelian randomization (MR) results for 30 significant polygenic score (PGS)-based associations.
Forest plots show associations between female health conditions as outcomes and cardiometabolic phenotypes as exposures in a 1-sample (**A**) and 2-sample (**B**) MR analyses. Genetic instruments are the cardiometabolic PGSs in the 1-sample MR and genome-wide significant single nucleotide polymorphisms from genome-wide association studies in the 2-sample MR. **A**, Each point refers to β outcome per SD exposure for the body mass index (BMI) PGS (PGS_{BMI}) and log odds ratio (OR) outcome for all other exposures for each test performed as separated by ancestry. **B**, Each point refers to β outcome/SD exposure for BMI and β outcome/SD log OR exposure for all other variables across all methods used in sensitivity analyses for 2-sample MR. *P* values are reported in the last column in both panels. AFR indicates African ancestry; CAD, coronary artery disease; DBP, diastolic blood pressure; EUR, European ancestry; HT, Hypertension; IVW, inverse variance weighted; PP; pulse pressure; SBP, systolic blood pressure; and T2D, type 2 diabetes.

$P<0.05$), and many also remained FDR significant. In all relationships, the direction of the estimated causal effect aligned with the direction of effect in the PGS association analysis. The MR results suggested potential causal relationships between many of the cardiometabolic phenotypes and female health conditions, with the most significant associations being between T2D and gestational diabetes (all: $P=1.5\times10^{-11}$, $\beta=1.52$, SE=0.23; EUR: $P=4.2\times10^{-11}$, $\beta=1.7$, SE=0.26) and between BMI and PCOS (all: $P=1.1\times10^{-8}$, $\beta=0.0021$, SE=0.00037; EUR: $P=6.9\times10^{-8}$, $\beta=0.0021$, SE=0.00037).

We also performed 2-sample MR to leverage the power of larger GWAS of female-specific health conditions (Figure 3B). A few relationships passed FDR significance when using all 3 methods (Table S4), such as T2D and gestational diabetes (IVW: $\beta=0.58$, SE=0.034, $P=2.74\times10^{-65}$; MR Egger: $\beta=0.53$, SE=0.079, $P=4.03\times10^{-11}$; weighted median: $\beta=0.5$, SE=0.042, $P=9.62\times10^{-32}$). Some associations were FDR significant when using the IVW method but became less significant when using the MR Egger and weighted median methods, likely because of the limitations of power of these methods. Other associations were still at least nominally significant ($P<0.05$) by all 3 methods, such as that between CAD and breast cancer (IVW: $\beta=-0.059$, SE=0.021, $P=0.0046$; MR Egger: $\beta=-0.1$, SE=0.049, $P=0.04$; weighted median: $\beta=-0.064$, SE=0.028, $P=0.021$). Horizontal pleiotropy was only significantly present in our results for BMI and breast cancer (intercept=0.015, SE=0.0056, $P=0.0089$), BMI and gestational diabetes (intercept=0.35, SE=0.011, $P=0.0016$), and T2D and gestational hypertension (intercept=0.011, SE=0.0027, $P=9.61\times10^{-5}$). Many relationships with significant pleiotropy still showed at least nominally significant causal effects when accounting for pleiotropy through the MR Egger approach, such as BMI with breast cancer (MR Egger: $\beta=-0.92$, SE=0.18, $P=4.23\times10^{-6}$). When performing MR in the opposite direction, we found no relationships that were significant using all 3 methods (Table S5). Some associations were significant when using the IVW and weighted median methods but became insignificant when using MR Egger, such as preeclampsia and SBP (IVW: $\beta=3.46$, SE=0.65, $P=1.12\times10^{-7}$; MR Egger: $\beta=7.5$, SE=14.9, $P=0.7$; weighted median: $\beta=3.41$, SE=0.41, $P=6.06\times10^{-17}$) and gestational diabetes and T2D (IVW: $\beta=0.2$, SE=0.049, $P=2.98\times10^{-5}$; MR Egger: $\beta=0.13$, SE=0.11, $P=0.26$; weighted median: $\beta=0.12$, SE=0.011, $P=3\times10^{-31}$). Weak instrument strength could also bias the causal estimate of the exposure on the outcome and its level of significance. The mean F statistics of the SNPs used in each MR analysis were all at least above 57, values that suggest the SNPs included in our analyses were unlikely to be weak and thus do not bias our results (Table S6). Notably, significance and direction of effect remained relatively consistent for most relationships across 1- and 2-sample MR methods.

Role of Population Stratification

Population stratification can confound the results of PGS association and MR analyses. Therefore, we tested the PGS associations with PCs (Table S7). The high R^2 values for most cardiometabolic PGSs indicated that much of the variance in the PGSs 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 PGS 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, in our prior analyses we accounted for PCs in both the PGS association analyses, used likelihood ratio test to compare the null model with the PGS model, and adjusted with PCs in the 1-sample MR analyses. Additionally, we performed 1-sample MR without adjusting for covariates (PCs and birth year) and compared the results with those of the adjusted MR analysis (Table S8). Relationships were generally more significant when using the PGS without including covariates, evidence that further suggests the presence of population stratification in our data sets and the need to adjust for covariates in our models.

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 PGSs tended to have more health conditions than those with low-risk cardiometabolic PGSs (Figure 4 and Figures S15 through S19). In particular, pregnancy-related conditions were relatively more prevalent in participants <25 years of age with high PGSs than low PGSs for all cardiometabolic PGSs. In addition, participants with low blood pressure PGSs tended to have lower prevalence of gestational hypertension and preeclampsia than other PGSs in the younger age groups (<25 and 25–39 years of age). We observed similar trends for PGS_{T2D} and gestational diabetes and for PGS_{BMI} and PCOS. We also observed high prevalence of endometriosis and uterine fibrosis in participants with high-risk PGSs, particularly those between 25 and 39 years of age.

DISCUSSION

We observed genetic correlations between cardiometabolic traits and a variety of obstetric and gynecological

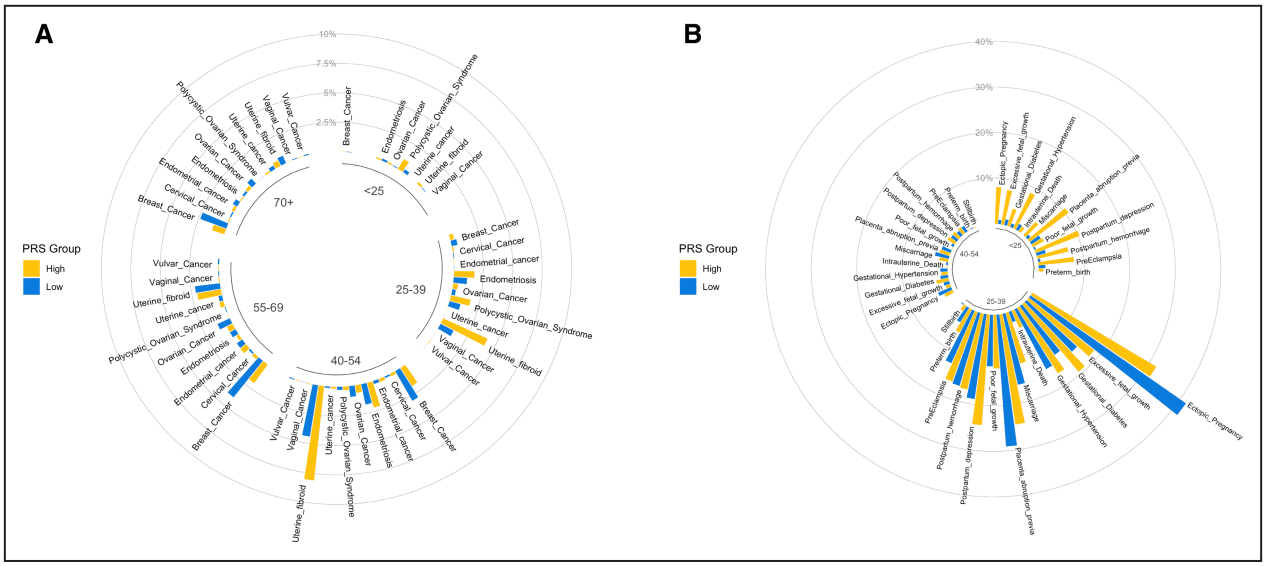


Figure 4. Chronology analyses of events from the electronic health records.
A circular plot shows disease prevalence among participants with high PGS_{BMI} (in yellow) and low PGS_{BMI} (in blue). General female health conditions are shown in (**A**), and pregnancy and childbirth-related phenotypes are shown in (**B**). The circular plots are divided into 5 age categories (<25, 25–39, 40–54, 55–69, and ≥70 years) for general female health conditions and 3 age categories (<25, 25–39, and 40–54 years) for pregnancy-related phenotypes. BMI indicates body mass index; PGS, polygenic scores; and PRS, polygenic risk score.

disorders, and these correlations suggest some significant degree of shared genetic cause among all the phenotypes tested in this study. We investigated the effects of this shared genetic cause by generating PGSs for cardiometabolic phenotypes and testing their associations with various female-specific health conditions. The PGSs were generally predictive of their corresponding cardiometabolic phenotypes as well as several female-specific health conditions. Prior research based mainly on nongenetic factors provided evidence for some of these associations, such those between BMI and PCOS and between T2D and gestational diabetes.[11,12] Epidemiological studies have also provided evidence of a positive correlation between obesity and endometrial cancer.[39] Our results suggest that certain relationships between cardiometabolic phenotypes and female-specific health conditions share a genetic basis. Furthermore, our MR results support that cardiometabolic phenotypes have potential causal effects on some female-specific health conditions.

PGS_{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[40]; however, our results indicated that a high genetic risk of CAD was protective against breast cancer, because participants with high PGS_{CAD} had lower incidence of breast cancer than participants with low PGS_{CAD}. When we considered only participants with high PGS_{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 PGS_{CAD} on breast cancer, suggest that more research should be undertaken to uncover the mechanisms behind this association.[41] Our MR results also suggest a weak but significant negative causal relationship between CAD and breast cancer, and the effect of horizontal pleiotropy for this relationship was also insignificant. In our analyses, we also did not consider breast cancer subtype, which displays differences in key characteristics such as gene expression, disease progression, and treatment response.[42] Breast cancer in our target data sets was defined using only *ICD-9 and ICD-10* codes and did not consider subtypes, and the breast cancer GWASs we used in this study also included cases of various subtypes. Further research incorporating breast cancer subtype could lead to interesting discoveries about the potential shared genetic factors for CAD and breast cancer.

The 3 blood pressure–related PGSs showed varying degrees of association with hypertensive diseases

during pregnancy across ancestry groups. PGS_{DBP} was significantly associated with gestational hypertension and preeclampsia in only participants of African ancestry, whereas only PGS_{SBP} was significantly associated with gestational hypertension and preeclampsia in all participants and participants of European and African ancestry. Prior studies also identified variability in predictive performance when different blood pressure measurements were used to predict hypertension and other diseases. Biases in collecting data from EHRs may have confounded the significance of these associations, and further studies are needed 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 associate with higher prevalence of certain health conditions at a young age, even in the absence of an overall association. For example, PGS_{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 PGS_{BMI} than in participants with low PGS_{BMI}. Similarly, there was no overall association between PGS_{SBP} and PCOS or endometriosis, but participants who developed PCOS before 25 years of age or endometriosis at 25 to 40 years of age were more likely to have high PGS_{SBP} than low PGS_{SBP}. Patients with high genetic risk of cardiometabolic phenotypes may therefore also have an elevated risk of early development of female-specific health conditions.

Our study establishes a link between the genetic risks of cardiometabolic phenotypes and several diseases that are unique to women; however, we estimated the genetic risk using PGSs alone, which are based on common variants and do not include the effects of rare variants and copy-number variations. In addition, we did not consider clinical or environmental factors, such as education level and socioeconomic status. Several studies have shown the effect of nongenetic risk factors on disease risk.[45,46] We found that population stratification was potentially present in our cohorts, so we accounted for PCs in our analyses accordingly. However, our focus on PGSs reflects the limitations of incorporating multiple modalities into analyses. Current interest in building integrative risk models suggests that this gap can be closed in the near future.[20,47] Furthermore, the weaker performance of PGSs 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. We calculated PGS using PRS-CS with a fixed global shrinkage parameter to reduce computational time, and we chose to set this parameter to 0.01 as recommended by the authors of PRS-CS for GWASs of smaller sample sizes for highly polygenic traits. Testing different values for the shrinkage parameter may improve the association strength of our PGS and may more accurately identify associations to female-specific health conditions. The PGSs were also found to have bimodal distributions. In most cases, bimodal distributions are not expected in a population sample, and the bimodal shape could indicate biased sampling in our 2 cohorts for phenotypes associated with the PGSs such as cardiometabolic conditions.

The small sample size for some conditions may have limited the power of our 1-sample MR analysis. 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 tested for pleiotropy and included methods such as MR Egger that account for pleiotropy and other confounding factors but may not have captured all their effects. Lastly, we also assessed whether our MR analyses was robust to weak instrument bias. Large F statistics (>57) in our analyses reflected the strength of our instrument variable and suggest that the results of our analyses do not suffer from the weak instrument bias.

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. In addition, because we generated only cardiometabolic PGSs in this study, PGSs specifically calculated for female-specific health conditions would provide a better understanding on an individual's genetic risk for female-specific health conditions and the risk for early development of these conditions.

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. Future studies should incorporate PGSs with other genetic and nongenetic risk factors and study their effects on larger and more diverse multiancestry populations.

APPENDIX

Regeneron Genetics Center Banner Author List

RGC Management and Leadership Team

Goncalo Abecasis, DPhil; Aris Baras, MD; Michael Cantor, MD; Giovanni Coppola, MD; Andrew Deubler, Aris Economides, PhD; Katia Karalis, PhD; Luca A.

Lotta, MD, PhD; John D. Overton, PhD; Jeffrey G. Reid, PhD; Katherine Siminovitch, MD; Alan Shuldiner, MD.

Sequencing and Laboratory Operations

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

Clinical Informatics

Amelia Averitt, PhD; Nilanjana Banerjee, PhD; Michael Cantor, MD; Dadong Li, PhD; Sameer Malhotra, MD; Deepika Sharma, MHI; Jeffrey Staples, PhD.

Genome Informatics

Xiaodong Bai, PhD; Suganthi Balasubramanian, PhD; Suying Bao, PhD; Boris Boutkov, PhD; Siying Chen, PhD; Gisu Eom, BS; Lukas Habegger, PhD; Alicia Hawes, BS; Shareef Khalid; Olga Krasheninina, MS; Rouel Lanche, BS; Adam J. Mansfield, BA; Evan K. Maxwell, PhD; George Mitra, BA; Mona Nafde, MS; Sean O'Keeffe, PhD; Max Orelus, BA; Razvan Panea, PhD; Tommy Polanco, BA; Ayesha Rasool, MS; Jeffrey G. Reid, PhD; William Salerno, PhD; Jeffrey C. Staples, PhD; Kathie Sun, PhD.

Analytical Genomics and Data Science

Goncalo Abecasis, DPhil; Joshua Backman, PhD; Amy Damask, PhD; Lee Dobbyn, PhD; Manuel Allen Revez Ferreira, PhD; Arkopravo Ghosh, MS; Christopher Gillies, PhD; Lauren Gurski, BS; Eric Jorgenson, PhD; Hyun Min Kang, PhD; Michael Kessler, PhD; Jack Kosmicki, PhD; Alexander Li, PhD; Nan Lin, PhD; Daren Liu, MS; Adam Locke, PhD; Jonathan Marchini, PhD; Anthony Marcketta, MS; Joelle Mbatchou, PhD; Arden Moscati, PhD; Charles Paulding, PhD; Carlo Sidore, PhD; Eli Stahl, PhD; Kyoko Watanabe, PhD; Bin Ye, PhD; Blair Zhang, PhD; Andrey Ziyatdinov, PhD.

Therapeutic Area Genetics

Ariane Ayer, BS; Aysegul Guvenek, PhD; George Hindy, PhD; Giovanni Coppola, MD; Jan Freudenberg, MD; Jonas Bovijn, MD; Katherine Siminovitch, MD; Kavita Praveen, PhD; Luca A. Lotta, MD; Manav Kapoor, PhD; Mary Haas, PhD; Moeen Riaz, PhD; Niek Verweij, PhD; Olukayode Sosina, PhD; Parsa Akbari, PhD; Priyanka Nakka, PhD; Sahar Gelfman, PhD; Sujit Gokhale, BE; Tanima De, PhD; Veera Rajagopal, PhD; Alan Shuldiner, MD; Bin Ye, PhD; Gannie Tzoneva, PhD; Juan Rodriguez-Flores, PhD.

RGC Biology

Shek Man Chim, PhD; Valerio Donato, PhD; Aris Economides, PhD; Daniel Fernandez, MS; Giusy Della Gatta, PhD; Alessandro Di Gioia, PhD; Kristen Howell, MS; Katia Karalis, PhD; Lori Khrimian, PhD; Minhee Kim, PhD; Hector Martinez; Lawrence Miloscio, BS; Sheilyn Nunez, BS; Elias Pavlopoulos, PhD; Trikaldarshi Persaud, BS.

Research Program Management & Strategic Initiatives

Esteban Chen, MS; Marcus B. Jones, PhD; Michelle G. LeBlanc, PhD; Jason Mighty, PhD; Lyndon J. Mitnaul, PhD; Nirupama Nishtala, PhD; Nadia Rana, PhD.

ARTICLE INFORMATION

Received May 17, 2022; accepted November 11, 2022.

Affiliations

Graduate Program in Genomics and Computational Biology, University of PennsylvaniaPhiladelphia, PA (B.X.); Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt University Medical CenterNashville, TN (D.R.V.E.); Department of Genetics, University of PennsylvaniaPhiladelphia, PA (A.L., T.D., M.R.); Department of Obstetrics and Gynecology, Brigham and Women's HospitalBoston, MA (K.G.); Department of Surgery, University of PennsylvaniaPhiladelphia, PA (B.K.); Department of Biomedical Informatics, Columbia UniversityNew York, NY (C.W., N.E.); Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington Medical CenterSeattle, WA (G.P.J.); Center for Applied Genomics, Children's Hospital of PhiladelphiaPhiladelphia, PA (H.H.); Center for Autoimmune Genomics and Etiology and Division of Allergy & Immunology, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of CincinnatiOH (L.K.); Department of Biomedical Informatics, Vanderbilt University Medical CenterNashville, TN (W-Q.W.); Department of Preventive Medicine, Feinberg School of Medicine, Northwestern UniversityEvanston, IL (Y.L.); Department of Biostatistics and Epidemiology (D.K.)and Department of Pathology and Laboratory Medicine (S.S.V.).

Acknowledgments

The authors acknowledge PMBB for providing data and thank the patient participants of Penn Medicine who consented to participate in this research program. The authors also thank the PMBB team and Regeneron Genetics Center for providing genetic variant data for analysis. The PMBB is approved under institutional review board protocol number 813913 and is supported by the Perelman School of Medicine at the University of Pennsylvania. The authors also thank the authors of PRS-CS for providing us with the code needed to generate the multiancestry LD reference panel. Data set: SNPs and their weights used to calculate PGS for each cardiometabolic condition in eMERGE and PMBB data are available in supplementary material to download. B.X. performed the analysis. S.S.V. designed the study. B.X. and S.S.V. wrote the article. T.D. and A.L. helped with data extraction and analyses. H.H., L.K., N.E., W.-Q.W., and Y.L. helped with data generation. B.X., D.R.V.E., T.D., K.G., B.K., C.W., G.P.J., D.K., M.R., and S.S.V. critically reviewed and approved the article. Regeneron Genetics Center genotyped the Penn Medicine BioBank data set.

Sources of Funding

The eMERGE network was initiated and funded by National Human Genome Research Institute 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.

Supplemental Material

Tables S1–S8
Figures S1–S18

REFERENCES

1. Blüher M. Obesity: global epidemiology and pathogenesis. *Nat Rev Endocrinol*. 2019;15:288–298. doi: 10.1038/s41574-019-0176-8
2. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, Colagiuri S, Guariguata L, Motala AA, Ogurtsova K, 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. doi: 10.1016/j.diabres.2019.107843
3. 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. *Lancet*. 2016;388:1459–1544.
4. Gerdts E, Regitz-Zagrosek V. Sex differences in cardiometabolic disorders. *Nat Med*. 2019;25:1657–1666. doi: 10.1038/s41591-019-0643-8
5. Buddeke J, Bots ML, van Dis I, Visseren FL, Hollander M, Schellevis FG, Vaartjes I. Comorbidity in patients with cardiovascular disease in primary care: a cohort study with routine healthcare data. *Br J Gen Pract*. 2019;69:e398–e406. doi: 10.3399/bjgp19X702725
6. 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;24:45–50. doi: 10.1080/13814788.2017.1398318
7. 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;24:75–85. doi: 10.3122/jabfm.2011.01.100076
8. 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;8:1625. doi: 10.3390/jcm8101625
9. Garovic VD, August P. Preeclampsia and the future risk of hypertension: the pregnant evidence. *Curr Hypertens Rep*. 2013;15:114–121. doi: 10.1007/s11906-013-0329-4
10. 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;70:798–803. doi: 10.1161/HYPERTENSIONAHA.117.09246
11. Gambineri A, Pelusi C, Vicennati V, Pagotto U, Pasquali R. Obesity and the polycystic ovary syndrome. *Int J Obes*. 2002;26:883–896. doi: 10.1038/sj.ijo.0801994
12. Kwak SH, Jang HC, Park KS. Finding genetic risk factors of gestational diabetes. *Genomics Inform*. 2012;10:239–243. doi: 10.5808/GI.2012.10.4.239
13. Gambineri A, Patton L, Altieri P, Pagotto U, Pizzi C, Manzoli L, Pasquali R. Polycystic ovary syndrome is a risk factor for type 2 diabetes: results from a long-term prospective study. *Diabetes*. 2012;61:2369–2374. doi: 10.2337/db11-1360
14. Mu F, Rich-Edwards J, Rimm EB, Spiegelman D, Missmer SA. Endometriosis and risk of coronary heart disease. *Circ Cardiovasc Qual Outcomes*. 2016;9:257–264. doi: 10.1161/CIRCOUTCOMES.115.002224
15. 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. doi: 10.3332/ecancer.2018.803
16. Chang C, Zhang K, Veluchamy A, Hébert HL, Looker HC, Colhoun HM, Palmer CN, Meng W. A genome-wide association study provides new evidence that CACNA1C gene is associated with diabetic cataract. *Invest Ophthalmol Vis Sci*. 2016;57:2246–2250. doi: 10.1167/iovs.16-19332
17. Patel SR. Shared genetic risk factors for obstructive sleep apnea and obesity. *J Appl Physiol*. 1985;2005(99):1600–1606. doi: 10.1152/japplphysiol.00501.2005
18. 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;8:55915–55919. doi: 10.18632/oncotarget.19115
19. Rashkin SR, Graff RE, Kachuri L, Thai KK, Alexeeff SE, Blatchins MA, Cavazos TB, Corley DA, Emami NC, Hoffman JD, et al. Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. *Nat Commun*. 2020;11:4423. doi: 10.1038/s41467-020-18246-6
20. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med*. 2020;12:44. doi: 10.1186/s13073-020-00742-5
21. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50:1219–1224. doi: 10.1038/s41588-018-0183-z
22. Steinthorsdottir V, McGinnis R, Williams NO, Stefansdottir L, Thorleifsson G, Shooter S, Fadista J, Sigurdsson JK, Auro KM, Berezina G, et al. Genetic predisposition to hypertension is associated with preeclampsia in European and central Asian women. *Nat Commun*. 2020;11:5976. doi: 10.1038/s41467-020-19733-6
23. Fritsche LG, Gruber SB, Wu Z, Schmidt EM, Zawistowski M, Moser SE, Blanc VM, Brummett CM, Kheterpal S, Abecasis GR, 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;102:1048–1061. doi: 10.1016/j.ajhg.2018.04.001
24. Shen X, Howard DM, Adams MJ, Hill WD, Clarke TK, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Deary IJ, Whalley HC, AM MI. A phenome-wide association and Mendelian randomisation study of polygenic risk for depression in UK biobank. *Nat Commun*. 2020;11:2301. doi: 10.1038/s41467-020-16022-0
25. Stanaway IB, Hall TO, Rosenthal EA, Palmer M, Naranbhai V, Knevel R, Namjou-Khales B, Carroll RJ, Kiryluk K, Gordon AS, et al. The eMERGE genotype set of 83,717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype. *Genet Epidemiol*. 2019;43:63–81. doi: 10.1002/gepi.22167
26. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, Huffman JE, Assimes TL, Lorenz K, Zhu X, 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;52:680–691.
27. Justice AE, Winkler TW, Feitosa MF, Graff M, Fisher VA, Young K, Barata L, Deng X, Czajkowski J, Hadley D, 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.
28. Giri A, Hellwege JN, Keaton JM, Park J, Qiu C, Warren HR, Torstenson ES, Kovesdy CP, Sun YV, Wilson OD, et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat Genetics*. 2019;51:51–62.
29. van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res*. 2018;122:433–443.
30. 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. doi: 10.1186/s13742-015-0047-8
31. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N, Daly MJ, Price AL, Neale BM. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47:291–295. doi: 10.1038/ng.3211
32. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467:52–58.

33. 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;51:584–591. doi: 10.1038/s41588-019-0379-x
34. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, Peterson R, Domingue B. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun*. 2019;10:3328. doi: 10.1038/s41467-019-11112-0
35. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10:1776. doi: 10.1038/s41467-019-09718-5
36. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36:1–48. doi: 10.18637/jss.v036.i03
37. Pendergrass SA, Dudek SM, Crawford DC, Ritchie MD. Visually integrating and exploring high throughput phenome-wide association study (PheWAS) results using PheWAS-view. *BioData Mining*. 2012;5:5. doi: 10.1186/1756-0381-5-5
38. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, et al. The MR-base platform supports systematic causal inference across the human phenome. *elife*. 2018;7:e34408. doi: 10.7554/eLife.34408
39. Onstad MA, Schmandt RE, Lu KH. Addressing the role of obesity in endometrial cancer risk, prevention, and treatment. *J Clin Oncol*. 2016;34:4225–4230. doi: 10.1200/JCO.2016.69.4638
40. Mehta LS, Watson KE, Barac A, Beckie TM, Bittner V, Cruz-Flores S, Dent S, Kondapalli L, Ky B, Okwuosa T, et al. Cardiovascular disease and breast cancer: where these entities intersect: a scientific statement from the American Heart Association. *Circulation*. 2018;137:e30–e66. doi: 10.1161/CIR.0000000000000556
41. Clarke SL, Parham M, Lankester J, Shadyab AH, Liu S, Kooperberg C, Manson JE, Tcheandjieu C, Assimes TL. Broad clinical manifestations of polygenic risk for coronary artery disease in the women's health initiative. *Commun Med*. 2022;2:108. doi: 10.1038/s43856-022-00171-y
42. Johnson KS, Conant EF, Soo MS. Molecular subtypes of breast cancer: a review for breast radiologists. *J Breast Imaging*. 2021;3:12–24. doi: 10.1093/jbi/wbaa110
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*. 2017;19:603–610. doi: 10.1111/jch.13015
44. Sesso HD, Stampfer MJ, Rosner B, Hennekens CH, Gaziano JM, Manson JE, Glynn RJ. Systolic and diastolic blood pressure, pulse pressure, and mean arterial pressure as predictors of cardiovascular disease risk in men. *Hypertension*. 2000;36:801–807. doi: 10.1161/01.HYP.36.5.801
45. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet*. 2016;17:392–406. doi: 10.1038/nrg.2016.27
46. Dimovski K, Orho-Melander M, Drake I. A favorable lifestyle lowers the risk of coronary artery disease consistently across strata of nonmodifiable risk factors in a population-based cohort. *BMC Public Health*. 2019;19:1575. doi: 10.1186/s12889-019-7948-x
47. Kachuri L, Graff RE, Smith-Byrne K, Meyers TJ, Rashkin SR, Ziv E, Witte JS, Johansson M. Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. *Nat Commun*. 2020;11:6084. doi: 10.1038/s41467-020-19600-4

SUPPLEMENTAL MATERIAL

Table S1. ICD codes used to determine cases and controls for cardiometabolic and female health phenotypes

Phenotype	Cases	Controls
Cardiometabolic Phenotypes		
Coronary Artery Disease	414.0*, I25.1*	Females without codes from cases
Hypertension	401*, I10*	Females without codes from cases
Type 2 Diabetes	250*, E11*	Females without codes from cases
Women's Health Phenotypes		
Placenta abruption, previa	O44.*, O45.*, 641.0, 641.1, 641.2, 667, 762.0, 762.1, 762.2	Z32.01, O80.*, 650, V72.42 (ICD-9)but no code 632.* or O02.* and no codes from cases
Gestational Hypertension	O13.*, 642.4, 642.5, 642.6, 642.7	Z32.01, O80.*, 650, V72.42 (ICD-9), O09.*, V23.*(ICD-9), Z33.1, V22.2 (ICD-9)but no code 632.* or O02.* and no codes from cases
PreEclampsia	O14.*, O11.*, 642.4, 642.5, 642.6, 642.7	Z32.01, O80.*, 650, V72.42 (ICD-9), O09.*, V23.*(ICD-9), Z33.1, V22.2 (ICD-9)but no code 632.* or O02.* and no codes from cases
Poor fetal growth	O36.5*	Females with codes O80.*, 650.* but not the code in cases
Excessive fetal growth	036.6*, 656.6, 656.60, 656.63, 656.61	Females with codes O80.*, 650.* but not the code in cases
Intrauterine Death	656.4, 656.41, 656.43, O31.2, O36.4. O36.4XX0,O36.4XX1, O36.4XX2, O36.4XX3, O36.4XX4, O36.4XX5, O36.4XX9, 656.4,656.41.656.43 but not with any of codes O35.9XX0, Q95.0, Q95.1, Q95.8,Q95.9	Females with codes O80.*, 650.* but not the code in cases
Stillbirth	P95, Z37.1, V27.1 (ICD-10), V27.4 (ICD-10), Z37.4, V27.7(ICD-10), V35 (ICD-9) Z37.7 but not with any of codes O35.9XX0, Q95.0, Q95.1, Q95.8,Q95.9	Females with codes O80.*, 650.* but not the code in cases also no code Z35.208

Gestational Diabetes	O24.*, 648.8*	Z32.01, O80.*, 650, V72.42 (ICD-9), O09.*, V23.*(ICD-9), Z33.1, V22.2 (ICD-9) but no code 632.* or O02.* and no codes from cases
Ectopic Pregnancy	633.*, O00.*, V32.42(ICD-9), 761.4, O08.105,O08.806, O08.006, O008.104, O009, P01.4, CPT-codes: 59130, 59140, 59135, 59136, 59120, 59121	Z32.01, O80.*, 650, V72.42 (ICD-9), O09.*, V23.*(ICD-9), Z33.1, V22.2 (ICD-9)but no code 632.* or O02.* and no codes from cases and no codes O09.1, O09.11, O09.12, O09.13, O09.10
Miscarriage	632,O02.1,O31.10X0 but not with any of codes O35.9XX0, Q95.0, Q95.1, Q95.8,Q95.9	Z32.01, O80.*, 650, V72.42 (ICD-9), O09.*, V23.*(ICD-9), Z33.1, V22.2 (ICD-9)but no codes from cases
Uterine fibroid	D25.9, 218.9	Females without codes from cases
Endometriosis	N80.*, 617.*	Females without codes from cases
Polycystic Ovarian Syndrome	E28.2, 256.4	Females without codes from cases
Breast Cancer	C50.*, 174,175,233.0	Females without codes from cases
Vulvar Cancer	C51.*, 184.4	Females without codes from cases
Vaginal Cancer	C52.*, 184.0	Females without codes from cases
Cervical Cancer	C53.*, 180.*	Females without codes from cases
Endometrial cancer	C54.*, 182.*	Females without codes from cases
Uterine cancer	C55.*, 179	Females without codes from cases
Ovarian Cancer	C56.*, 183,186,220,220.0	Females without codes from cases
Postpartum depression	F53.*, O90.6, 648.4*	Females with codes O80.*, 650.* but not the code in cases
Preterm birth	O60.1*, 644.765	Females with codes O80.*, 650.* but not the code in cases
Postpartum hemorrhage	O72.*, 666.0,666.1,666.2	Females with codes O80.*, 650.* but not the code in cases

Table S2. Publicly available female health condition GWAS used in two-sample MR

Phenotype	Ancestry	Source	Sample Size (N cases)	Reference or PMID
Breast Cancer	EUR	BCAC	33832 (15748)	25751625
Endometrial Cancer	EUR	ECAC, E2C2, UKBB	121885 (12906)	30093612
Gestational Diabetes	EUR	FinnGen	116363 (6033)	https://r5.finngen.fi/
Gestational Hypertension	EUR	FinnGen	118990 (4255)	https://r5.finngen.fi/
Preeclampsia	EUR	FinnGen	123579 (8844)	https://r5.finngen.fi/
Polycystic Ovarian Syndrome	EUR	FinnGen	118870 (642)	https://r5.finngen.fi/
Postpartum Depression	EUR	FinnGen	67205 (7604)	https://r5.finngen.fi/

Table S3. PGS performance on cardiometabolic phenotypes in PMBB using LD panels of different ancestries

PRS	LD Panel	R^2/AUC in PMBB All	R^2/AUC in PMBB EUR	R^2/AUC in PMBB AFR
BMI	Multi	0.1542	0.07977	0.03556
	EUR	0.1528	0.7888	0.3381
	AFR	0.1538	0.07575	0.03663
CAD	Multi	0.8235	0.8303	0.799
	EUR	0.8237	0.8303	0.7992
	AFR	0.8238	0.8303	0.7994
DBP	Multi	0.0425	0.01555	0.04545
	EUR	0.0412	0.01425	0.04429
	AFR	0.0427	0.01564	0.04597
PP	Multi	0.8125	0.7797	0.839
	EUR	0.8123	0.7796	0.8387
	AFR	0.8124	0.7795	0.839
SBP	Multi	0.8144	0.7817	0.8416
	EUR	0.8143	0.7818	0.8413
	AFR	0.8144	0.7815	0.8415
T2D	Multi	0.754	0.7133	0.7366
	EUR	0.7544	0.7148	0.7364
	AFR	0.7529	0.7089	0.7369

Table S4. Direct tests for pleiotropy using MR Egger for two-sample Mendelian randomization

Exposure	Outcome	Intercept	SE	P value
BMI	Breast Cancer	0.015	0.0056	0.0089
	Endometrial Cancer	0.0033	0.0064	0.6
	Gestational Diabetes	0.035	0.011	0.0016
	Polycystic Ovarian Syndrome	-0.0048	0.03	0.88
CAD	Breast Cancer	0.0025	0.0026	0.34
PP	Gestational Hypertension	0.009	0.0094	0.338
	Preeclampsia	0.0034	0.0075	0.66
SBP	Gestational Hypertension	-7.86E-05	0.013	0.99
	Preeclampsia	-0.002	0.0093	0.83
T2D	Breast Cancer	0.00035	0.0016	0.83
	Gestational Diabetes	0.0021	0.0033	0.53
	Gestational Hypertension	0.011	0.0027	9.61E-05
	Polycystic Ovarian Syndrome	0.01	0.007	0.14
	Postpartum Depression	0.0037	0.0021	0.082

Table S5. Two-sample Mendelian randomization using female-specific health conditions as exposures and cardiometabolic phenotypes as outcomes

Exposure	Outcome	Method	Beta	SE	P value
Breast Cancer	BMI	IVW	-0.024	0.023	0.3
		Weighted median	-0.01	0.012	0.39
		MR Egger	0.028	0.06	0.64
	CAD	IVW	0.0095	0.01	0.34
		Weighted median	0.014	0.013	0.27
		MR Egger	0.0084	0.023	0.72
	T2D	IVW	0.0023	0.17	0.89
		Weighted median	0.0011	0.009	0.91
		MR Egger	0.026	0.039	0.51
Endometrial Cancer	BMI	IVW	0.034	0.022	0.13
		Weighted median	0.03	0.018	0.1
		MR Egger	0.021	0.12	0.87
Gestational Diabetes	BMI	IVW	-0.021	0.022	0.37
		Weighted median	0.0096	0.01	0.35
		MR Egger	0.0037	0.0099	0.71
	T2D	IVW	0.2	0.049	3.00E-05
		Weighted median	0.12	0.011	3.00E-31
		MR Egger	0.13	0.11	0.26
Polycystic Ovarian Syndrome	T2D	Wald ratio	-0.01	0.016	0.51
Preeclampsia	PP	IVW	0.91	0.27	0.00065
		Weighted median	0.84	0.25	0.00093
		MR Egger	5.75	4	0.39

	SBP	IVW	3.46	0.65	1.12E-07
		Weighted median	3.41	0.41	6.06E-17
		MR Egger	7.5	14.9	0.7

*No SNPs were left after filtering for genome-wide significance and pruning for the postpartum depression GWAS. PCOS had no overlapping SNPs with the BMI GWAS. SNPs could not be extracted for the gestational hypertension GWAS.
**Wald ratio was used for the PCOS and T2D MR since there was only one genome-wide significant SNP in the PCOS GWAS.

Table S6. Mean F-statistic of SNPs used in each two-sample MR analysis

Exposure	Outcome	Mean F-statistic
BMI	Breast Cancer	64.968
	Endometrial Cancer	64.968
	Gestational Diabetes	64.775
	Polycystic Ovarian Syndrome	64.775
CAD	Breast Cancer	67.841
PP	Gestational Hypertension	57.511
	Preeclampsia	57.511
SBP	Gestational Hypertension	57.848
	Preeclampsia	57.848
T2D	Breast Cancer	84.601
	Gestational Diabetes	83.931
	Gestational Hypertension	83.931
	Polycystic Ovarian Syndrome	83.931

Table S7. PRS PC associations in PMBB and eMERGE

PRS	Group	R^2
eMERGE		
BMI	All	0.503
	EUR	0.0727
	AFR	0.27
CAD	All	0.11
	EUR	0.0156
	AFR	0.0361
DBP	All	0.352
	EUR	0.135
	AFR	0.182
PP	All	0.346
	EUR	0.0469
	AFR	0.176
SBP	All	0.572
	EUR	0.271
	AFR	0.395
T2D	All	0.598
	EUR	0.174
	AFR	0.374
PMBB		
BMI	All	0.602
	EUR	0.00334
	AFR	0.227

CAD	All	0.163
	EUR	0.0106
	AFR	0.0158
DBP	All	0.451
	EUR	0.0423
	AFR	0.102
PP	All	0.517
	EUR	0.0117
	AFR	0.16
SBP	All	0.58
	EUR	0.0284
	AFR	0.198
T2D	All	0.647
	EUR	0.0428
	AFR	0.292

Table S8. One-sample Mendelian randomization using only PRS with no other covariates

Exposure	Outcome	Group	Beta	SE	P value
BMI	Breast Cancer	All	-0.00737	0.0027	0.00629
	Endometrial Cancer	All	6.94E-07	0.000402	0.999
		EUR	0.0014	0.000362	0.000112
	Gestational Diabetes	All	0.0097	0.00731	0.185
		EUR	0.0144	0.00373	0.000108
	Polycystic Ovarian Syndrome	All	0.00264	0.000678	9.79E-05
		EUR	0.00275	0.000387	1.18E-12
CAD	Breast Cancer	All	-0.427	0.0451	3.24E-21
		EUR	-0.2	0.0356	1.75E-08
	Postpartum Depression	EUR	2.62	1.61	0.104
HT (PP)	Gestational Hypertension	All	0.911	0.564	0.106
		EUR	1.73	0.671	0.00976
	Preeclampsia	All	0.915	0.369	0.0132
SBP	Gestational Hypertension	All	0.743	0.287	0.00978
		EUR	0.987	0.432	0.0223
		AFR	0.722	0.296	0.0149
	Preeclampsia	All	0.779	0.0176	9.35E-06
		EUR	0.637	0.165	0.000119
		AFR	0.853	0.314	0.00658
T2D	Breast Cancer	All	-0.182	0.0506	0.000315
		EUR	-0.114	0.0223	3.06E-07
	Gestational Diabetes	All	1.05	0.165	2.11E-10
		EUR	1.59	0.252	2.59E-10

	Gestational Hypertension	All	0.976	0.251	0.000102
		EUR	0.865	0.272	0.00146
	Polycystic Ovarian Syndrome	All	0.058	0.00717	5.67E-16
		EUR	0.0625	0.0108	5.95E-09
		AFR	0.0246	0.0244	0.312
	Postpartum Depression	All	1.33	0.77	0.0839

Figure S1. Genetic correlations between multi-ancestry cardiometabolic phenotypes and female-specific health conditions using European (left) and African (right) ancestry-derived GWASs

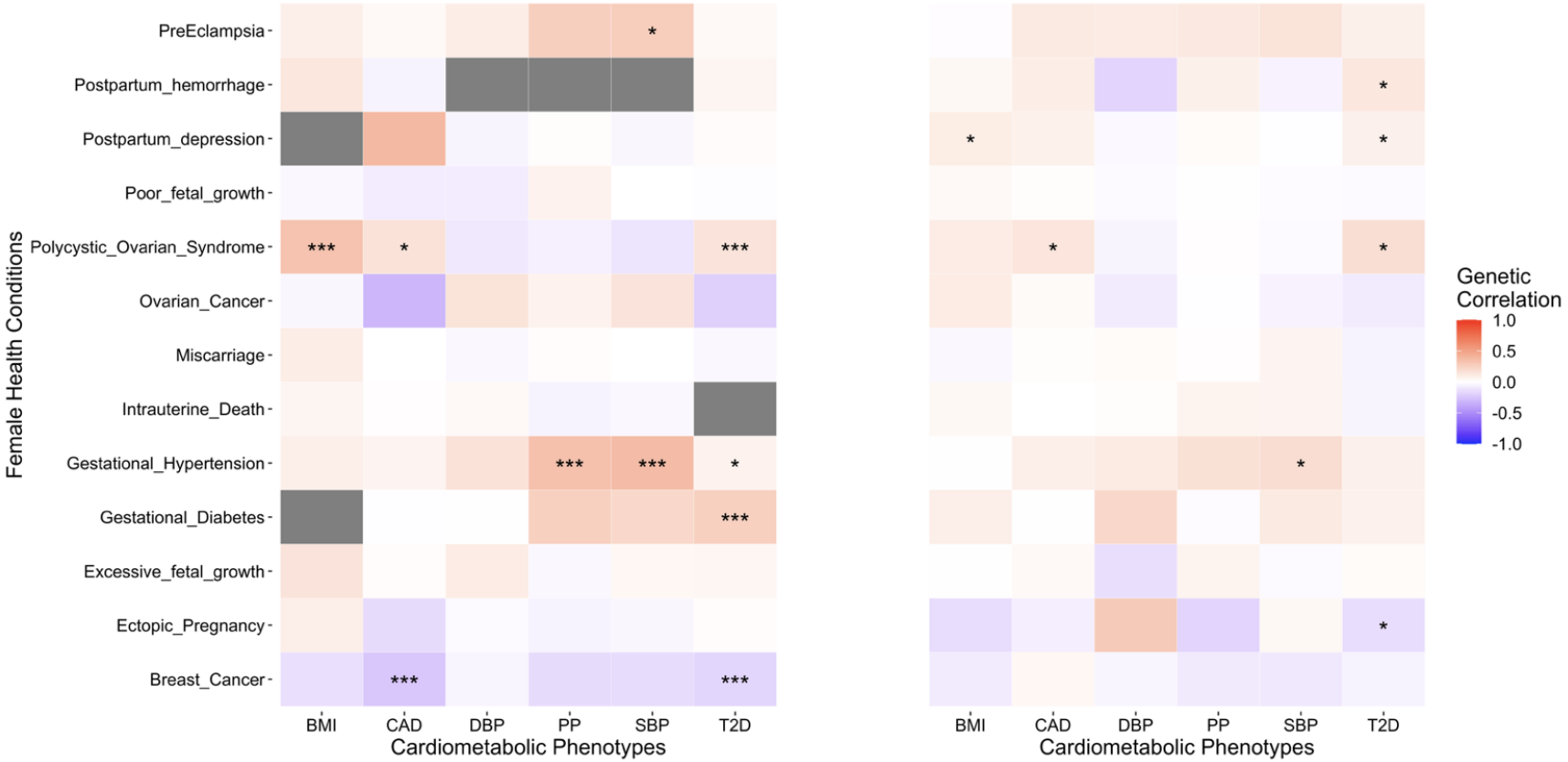


Figure S2. PRS_{BMI} distribution (eMERGE on left and PMBB on right)

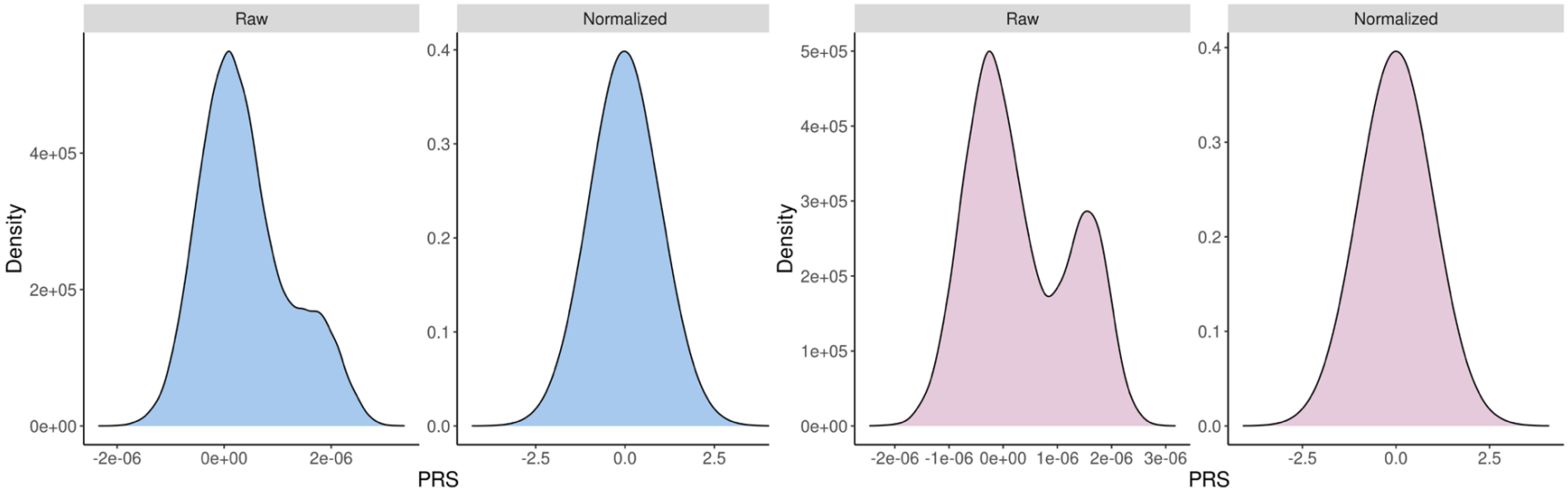


Figure S3. PRS_{CAD} distribution (eMERGE on left and PMBB on right)

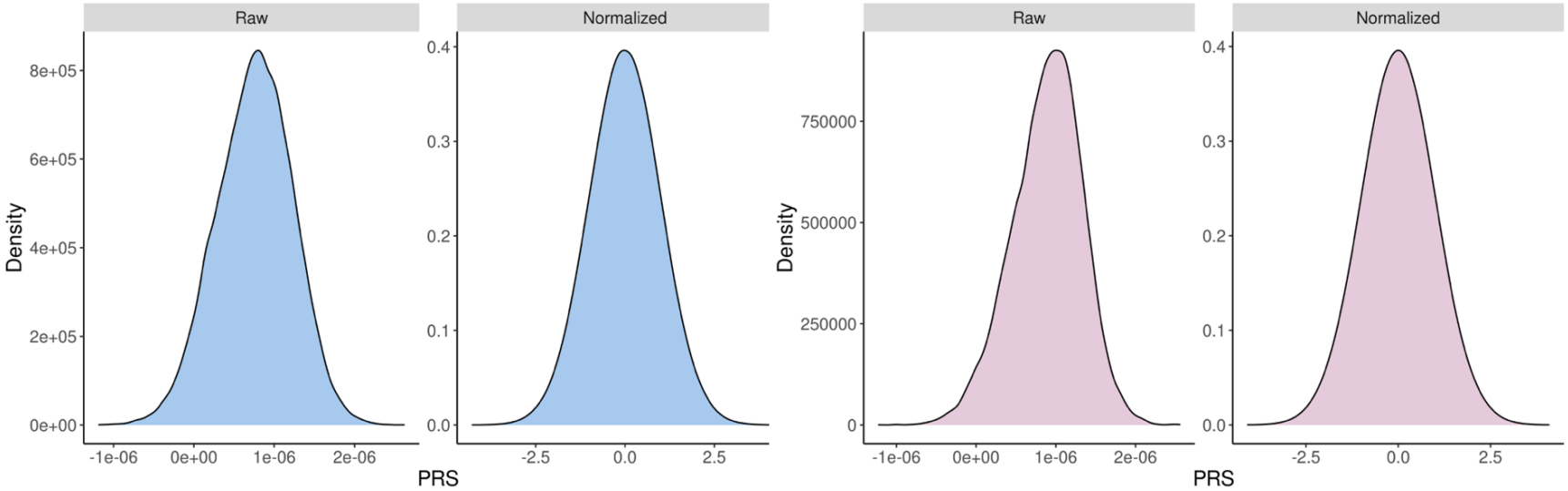


Figure S4. PRS_{DBP} distribution (eMERGE on left and PMBB on right)

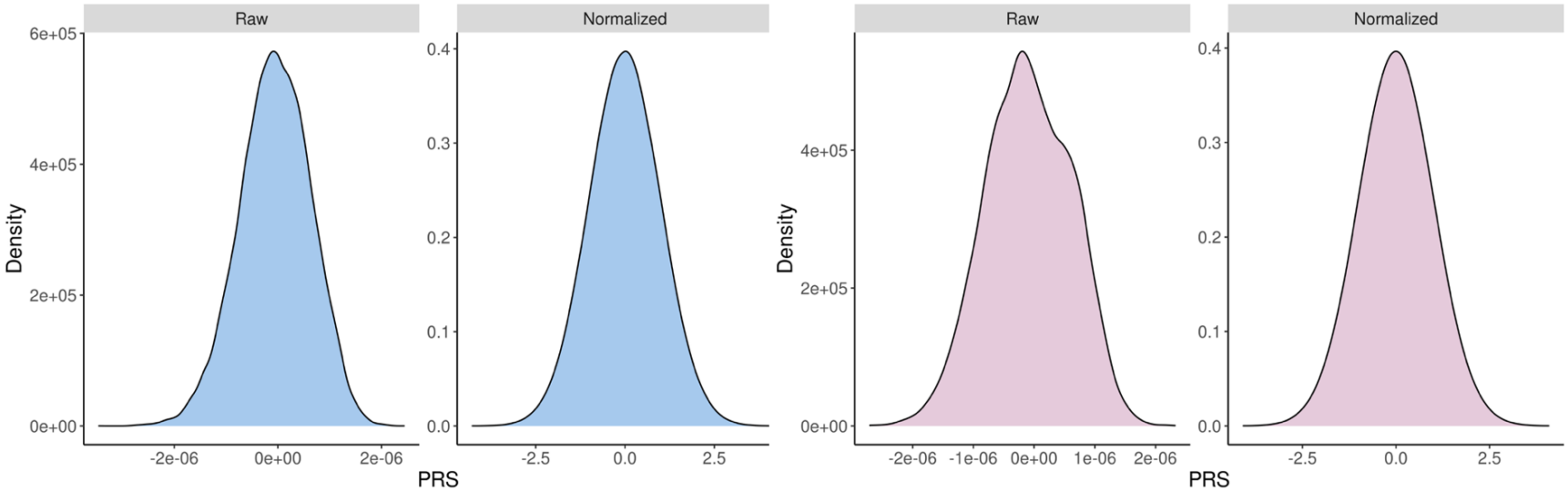


Figure S5. PRS_{PP} distribution (eMERGE on left and PMBB on right)

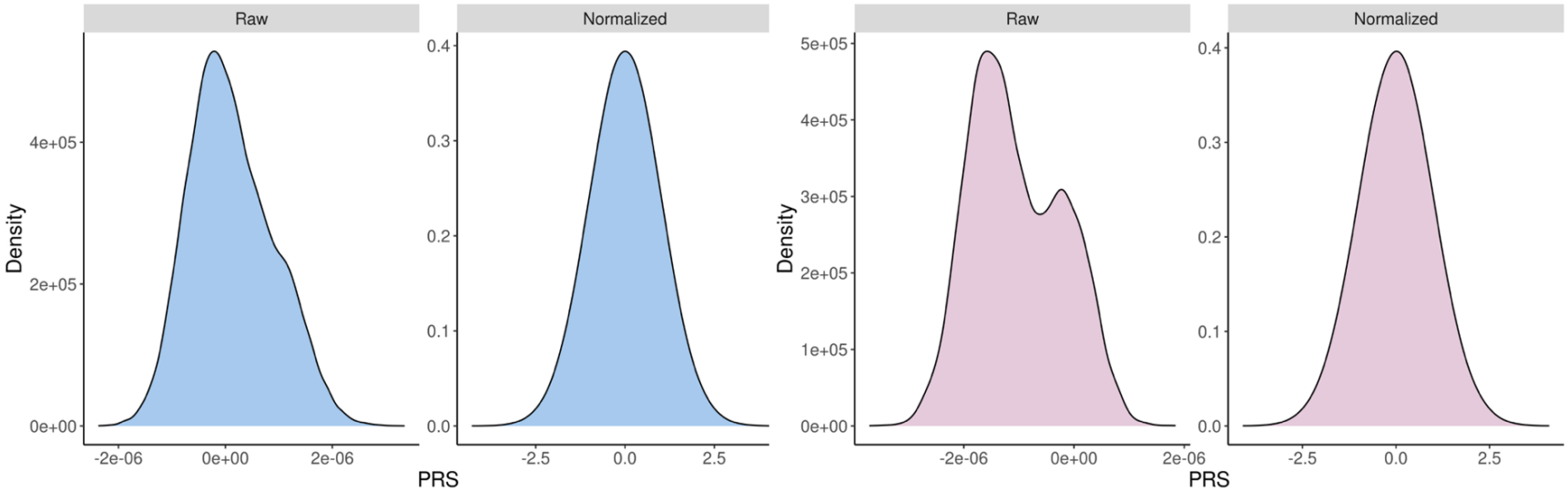


Figure S6. PRS_{SBP} distribution (eMERGE on left and PMBB on right)

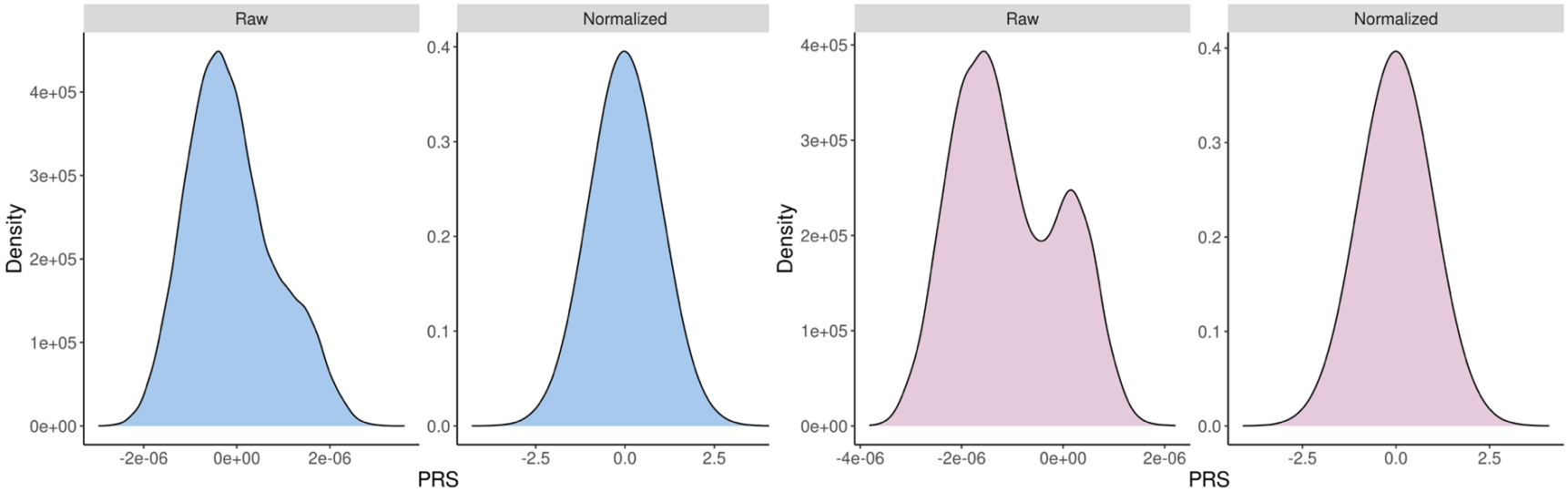


Figure S7. PRS_{T2D} distribution (eMERGE on left and PMBB on right)

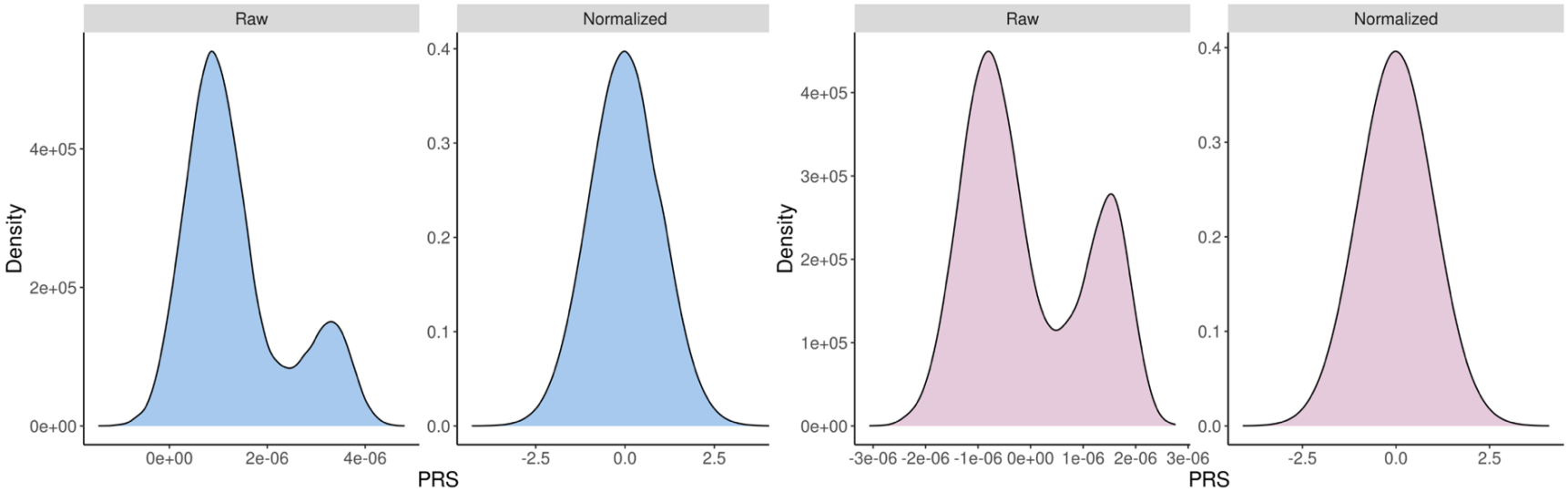


Figure S8. Prevalence by PRS quintile plot (PRS_{BMI} and breast cancer and PRS_{BMI} and endometrial cancer)

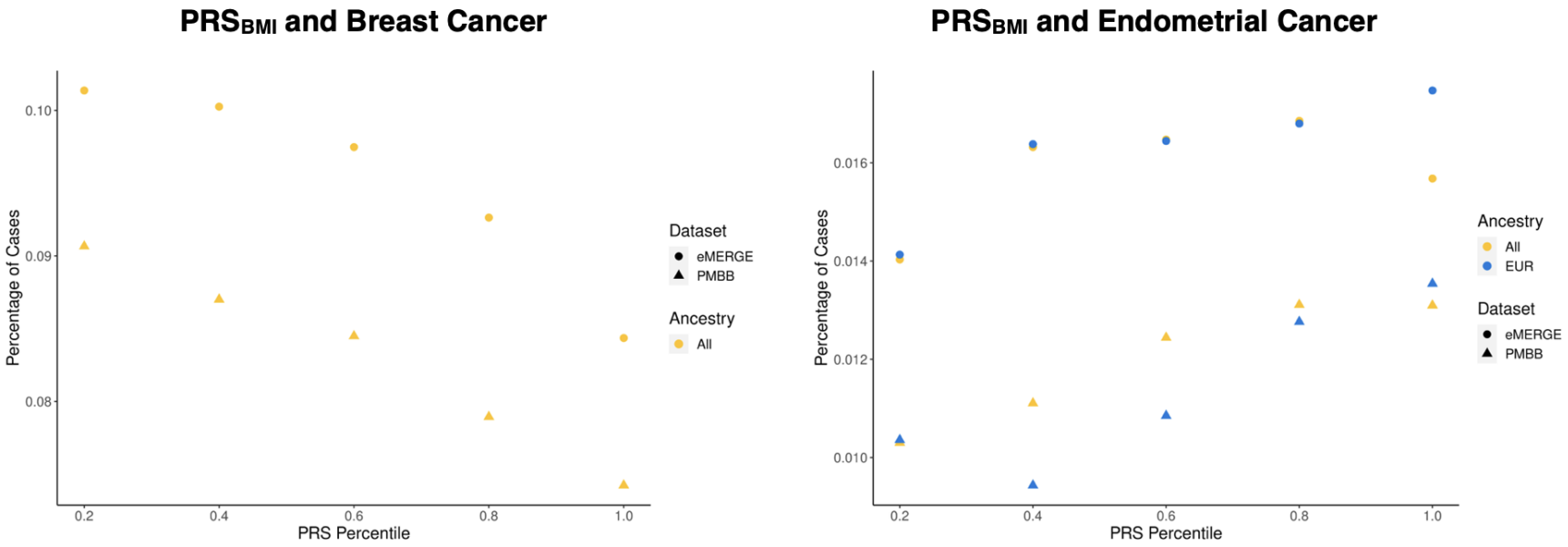


Figure S9. Prevalence by PRS quintile plot (PRS_{BMI} and gestational diabetes and PRS_{BMI} and PCOS)

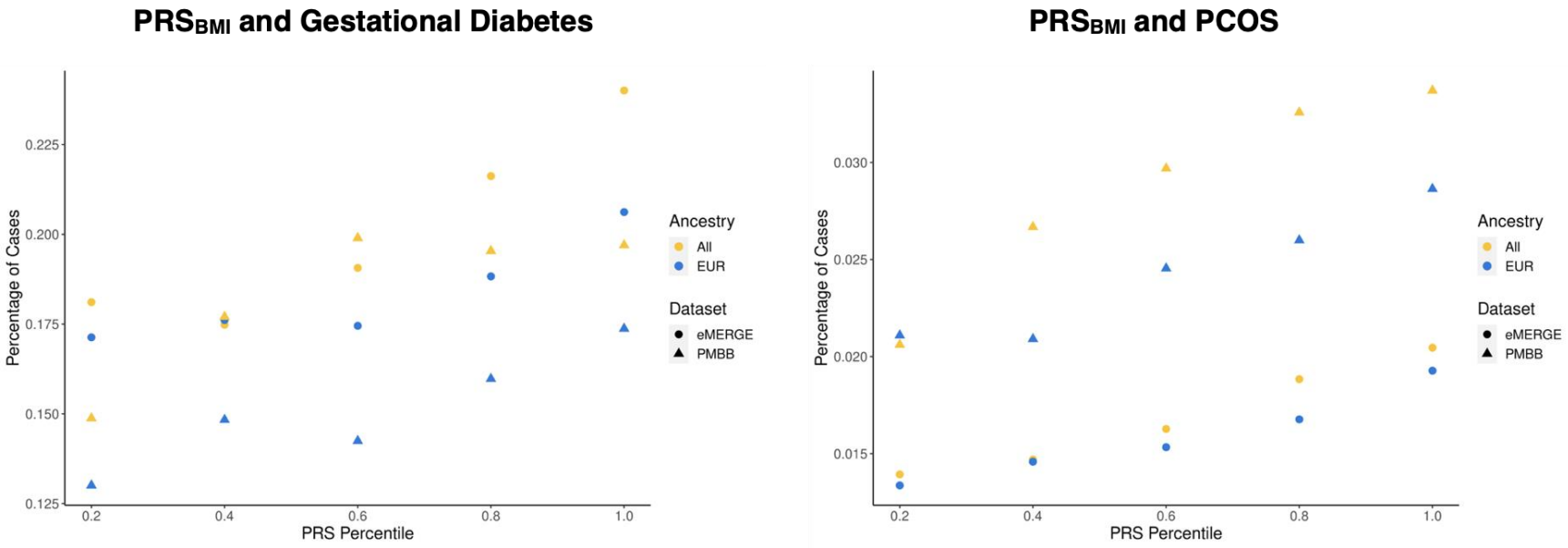


Figure S10. Prevalence by PRS quintile plot (PRS_{DBP} and excessive fetal growth and PRS_{DBP} and gestational hypertension)

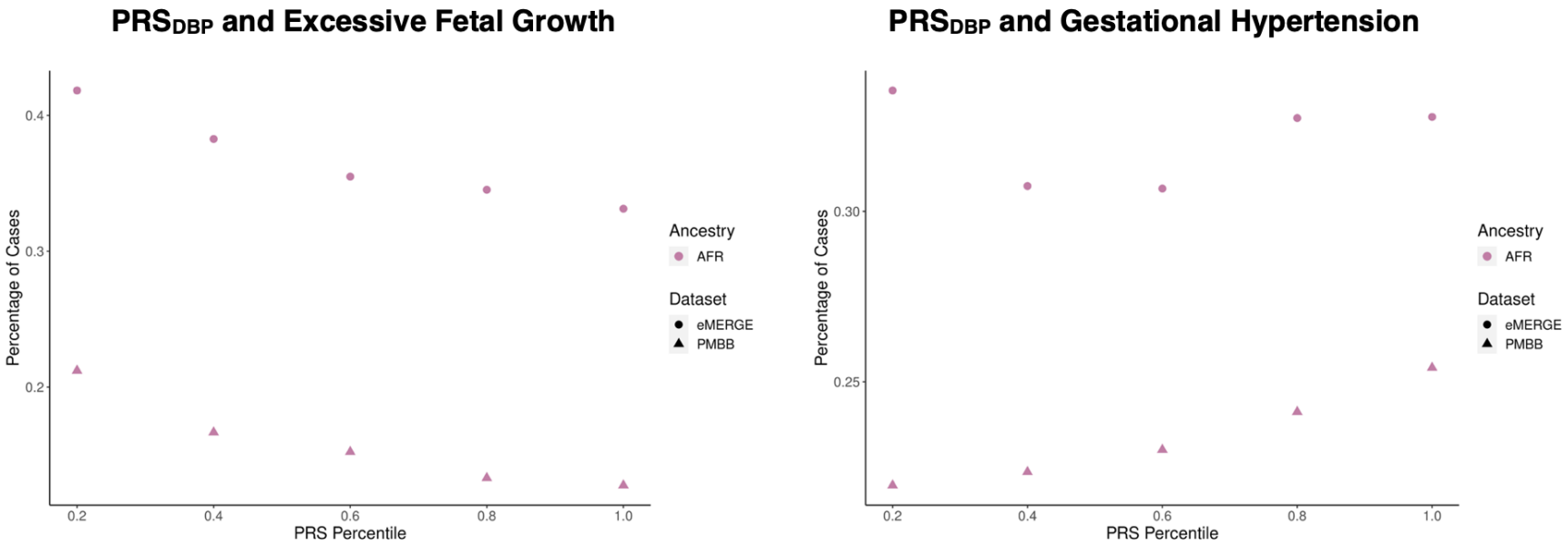


Figure S11. Prevalence by PRS quintile plot (PRS_{PP} and gestational hypertension and PRS_{PP} and preeclampsia)

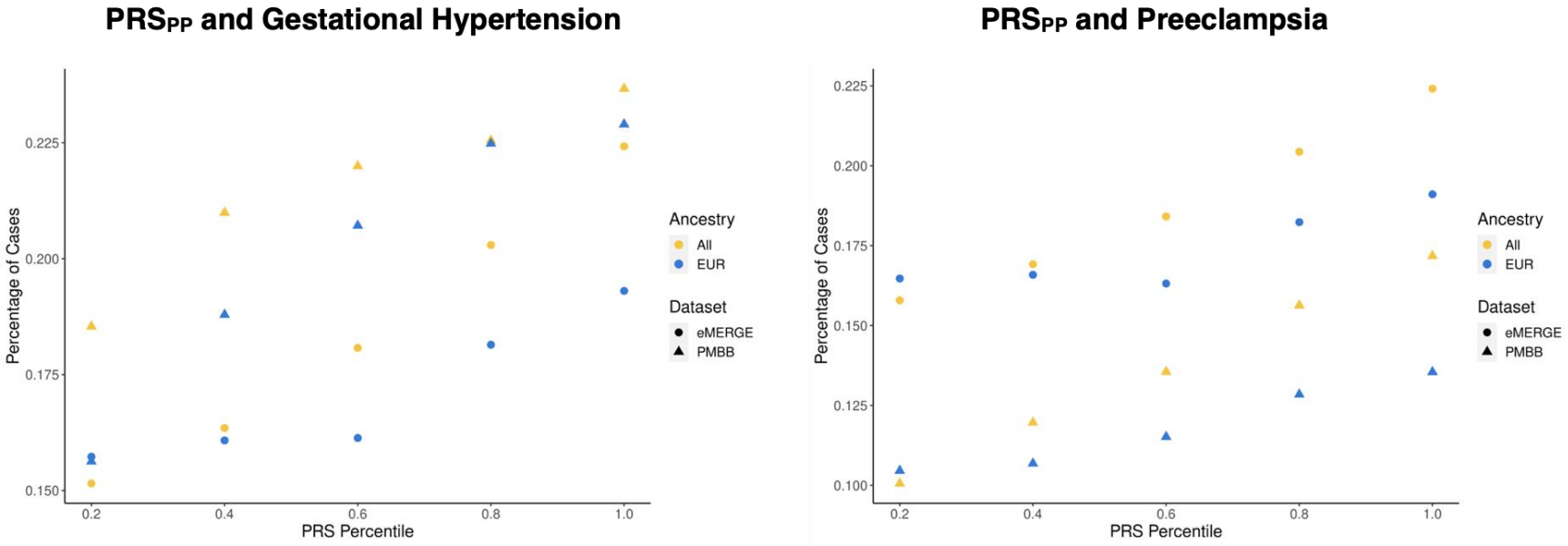


Figure S12. Prevalence by PRS quintile plot (PRS_{SBP} and gestational hypertension and PRS_{SBP} and preeclampsia)

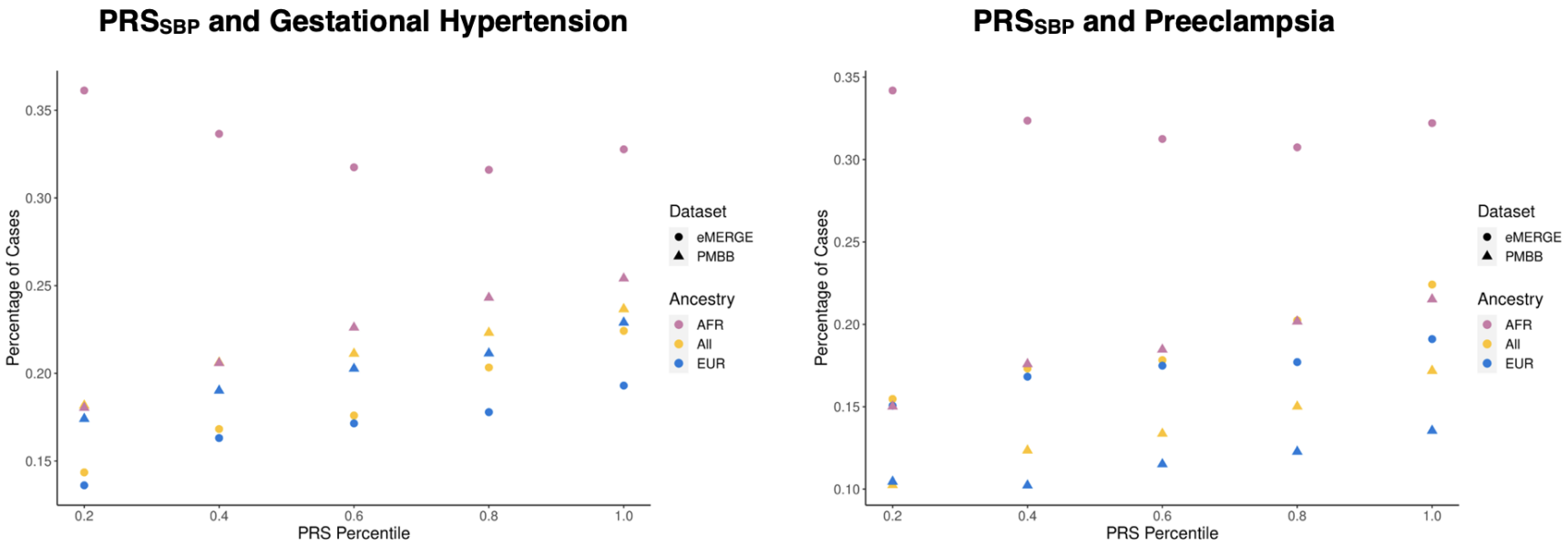


Figure S13. Prevalence by PRS quintile plot (PRS_{T2D} and breast cancer and PRS_{T2D} and gestational diabetes)

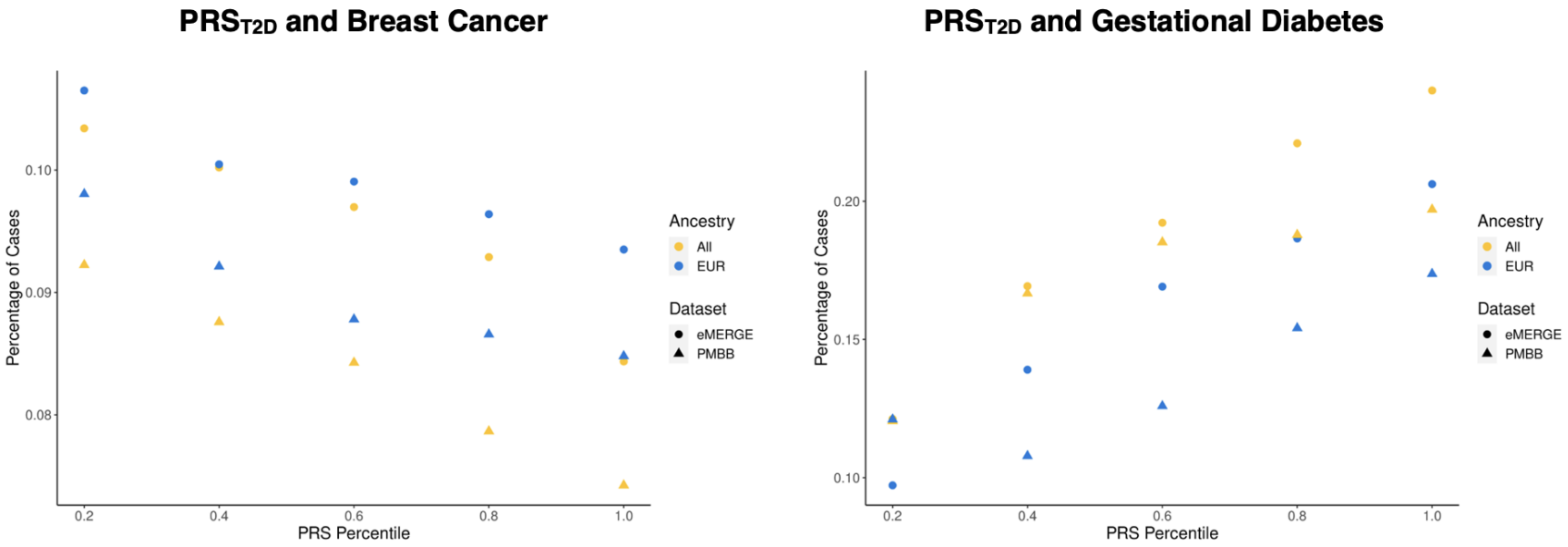


Figure S14. Prevalence by PRS quintile plot (PRS_{T2D} and gestational hypertension and PRS_{T2D} and PCOS)

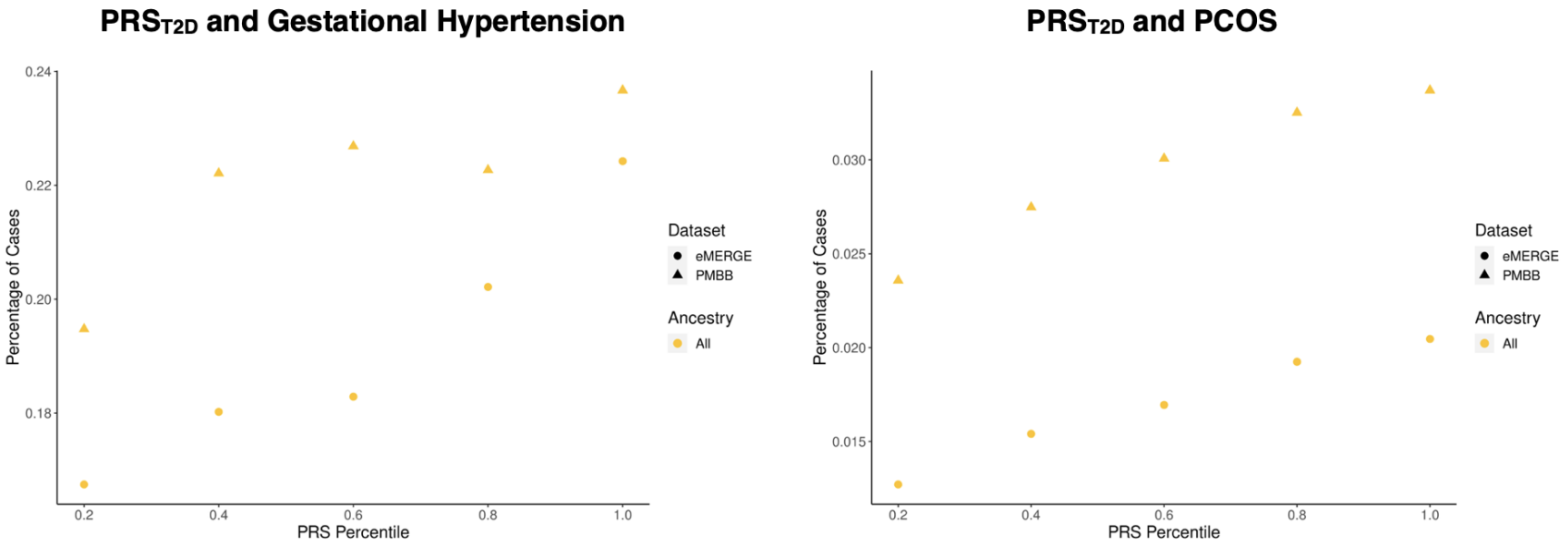


Figure S15. Chronological map for patients with high and low PRS_{CAD}

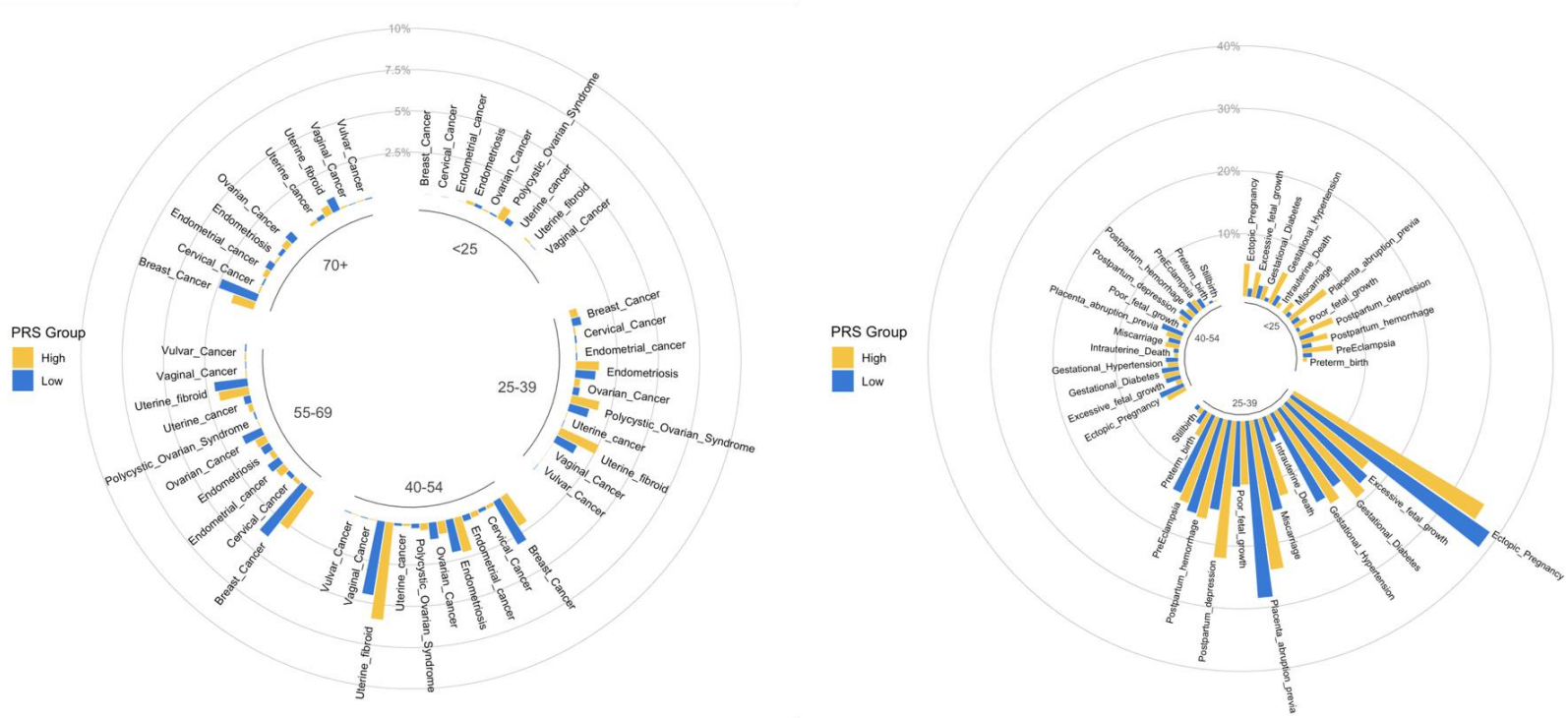


Figure S16. Chronological map for patients with high and low PRS_{DBP}

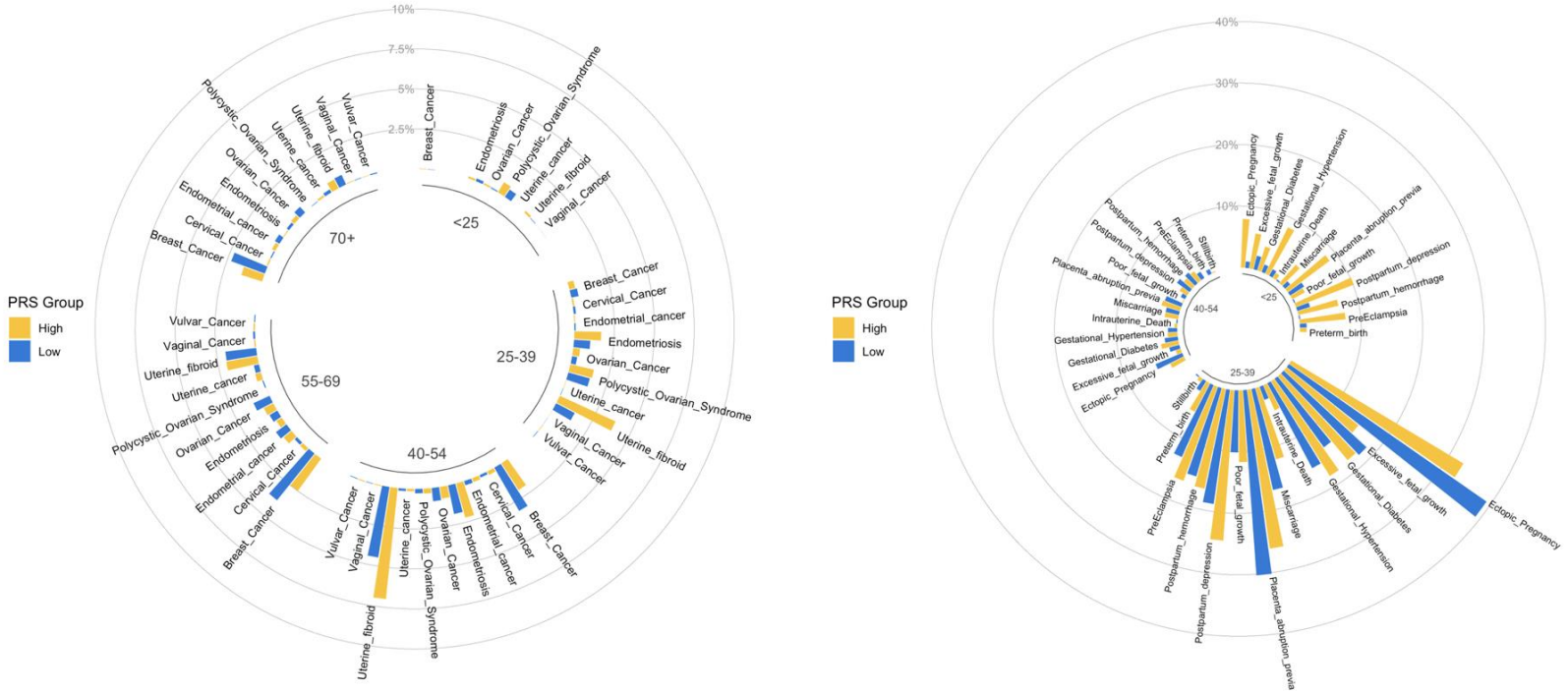


Figure S17. Chronological map for patients with high and low PRS_{PP}

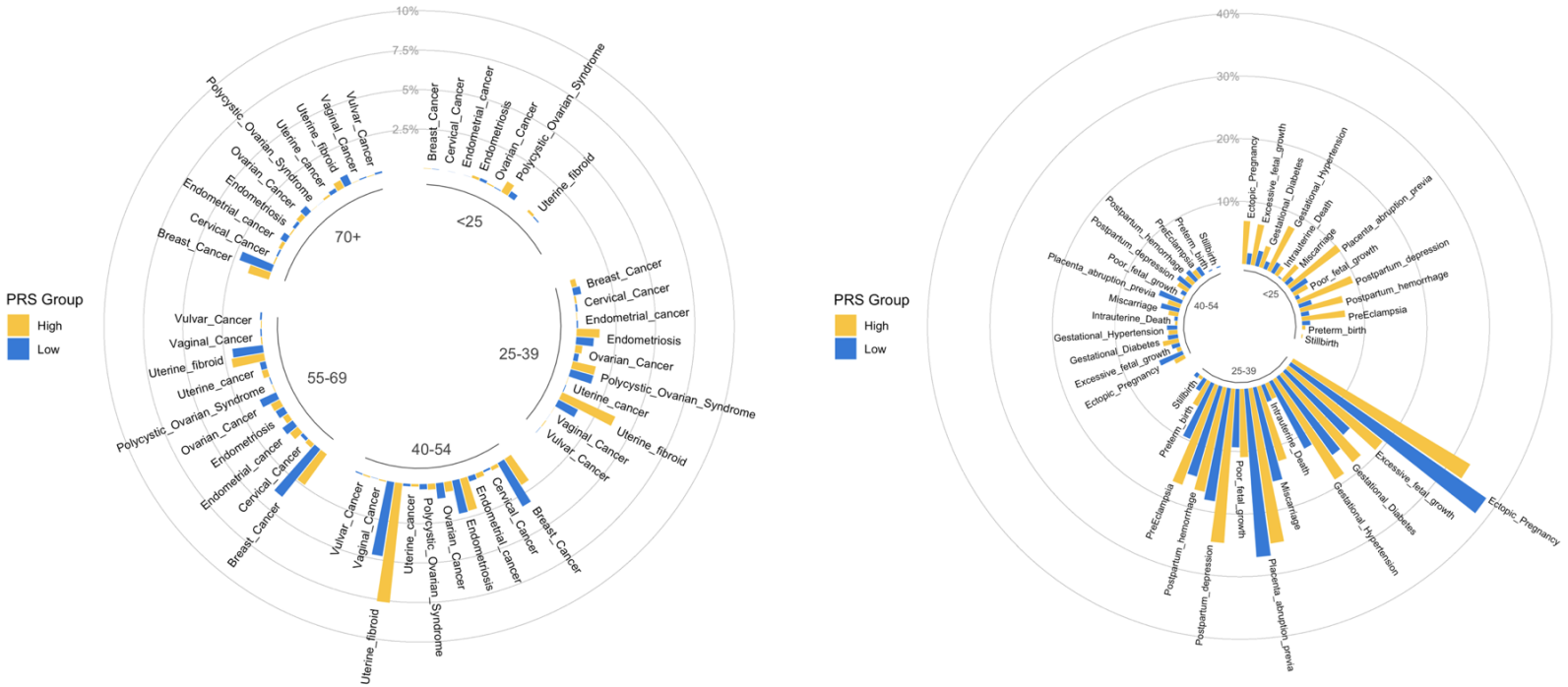
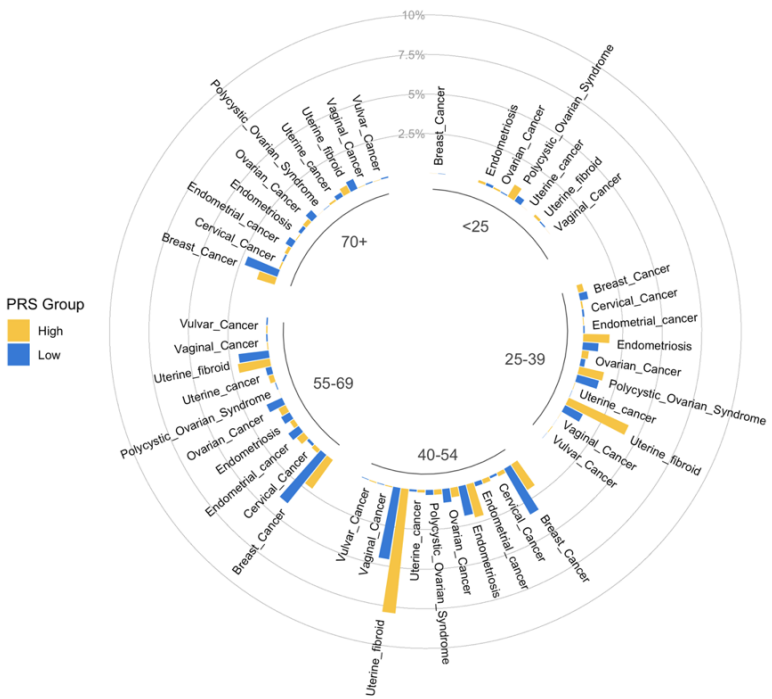


Figure S18. Chronological map for patients with high and low PRS_{SBP}



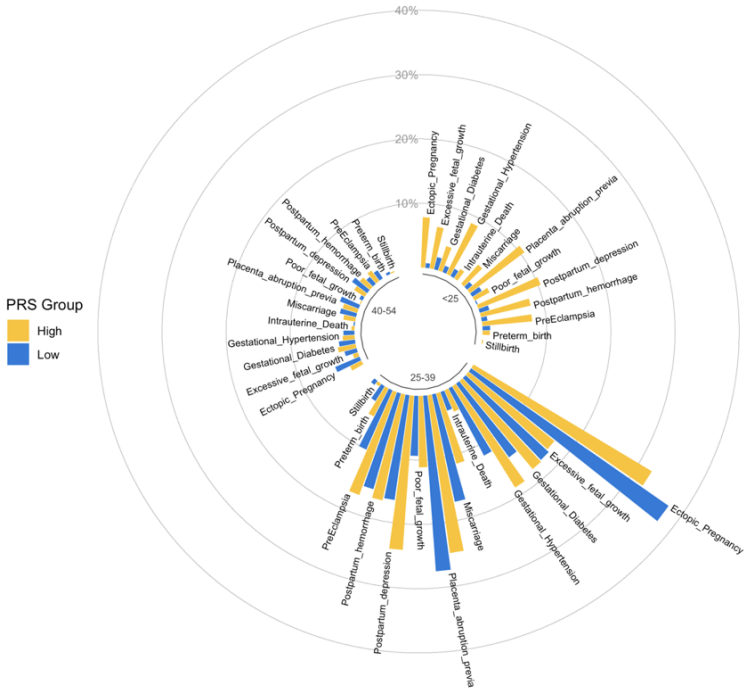


Figure S19. Chronological map for patients with high and low PRS_{T2D}