



# Genetic risk for major depressive disorder and loneliness in sex-specific associations with coronary artery disease

Jessica Dennis<sup>1,2</sup> · Julia Sealock<sup>1,2</sup> · Rebecca T. Levinson<sup>2,3</sup> · Eric Farber-Eger<sup>4</sup> · Jacob Franco<sup>5</sup> · Sarah Fong<sup>1,2</sup> · Peter Straub<sup>1,2</sup> · Donald Hucks<sup>1,2</sup> · Wen-Liang Song<sup>3</sup> · MacRae F. Linton<sup>3</sup> · Pierre Fontanillas<sup>6</sup> · Sarah L. Elson<sup>6</sup> · Douglas Ruderfer<sup>1,2</sup> · Abdel Abdellaoui<sup>7,8</sup> · Sandra Sanchez-Roige<sup>9</sup> · Abraham A. Palmer<sup>9,10</sup> · Dorret I. Boomsma<sup>11,12</sup> · Nancy J. Cox<sup>1,2</sup> · Guanhua Chen<sup>13</sup> · Jonathan D. Mosley<sup>14</sup> · Quinn S. Wells<sup>3</sup> · Lea K. Davis<sup>1,2</sup>

Received: 4 September 2019 / Revised: 24 September 2019 / Accepted: 18 November 2019 / Published online: 3 December 2019  
© The Author(s) 2019. This article is published with open access

## Abstract

Major depressive disorder (MDD) and loneliness are phenotypically and genetically correlated with coronary artery disease (CAD), but whether these associations are explained by pleiotropic genetic variants or shared comorbidities is unclear. To tease apart these scenarios, we first assessed the medical morbidity pattern associated with genetic risk factors for MDD and loneliness by conducting a phenome-wide association study in 18,385 European-ancestry individuals in the Vanderbilt University Medical Center biobank, BioVU. Polygenic scores for MDD and loneliness were developed for each person using previously published meta-GWAS summary statistics, and were tested for association with 882 clinical diagnoses ascertained via billing codes in electronic health records. We discovered strong associations with heart disease diagnoses, and next embarked on targeted analyses of CAD in 3893 cases and 4197 controls. We found odds ratios of 1.11 (95% CI, 1.04–1.18;  $P$   $8.43 \times 10^{-4}$ ) and 1.13 (95% CI, 1.07–1.20;  $P$   $4.51 \times 10^{-6}$ ) per 1-SD increase in the polygenic scores for MDD and loneliness, respectively. Results were similar in patients without psychiatric symptoms, and the increased risk persisted in females even after adjusting for multiple conventional risk factors and a polygenic score for CAD. In a final sensitivity analysis, we statistically adjusted for the genetic correlation between MDD and loneliness and re-computed polygenic scores. The polygenic score unique to loneliness remained associated with CAD (OR 1.09, 95% CI 1.03–1.15;  $P$  0.002), while the polygenic score unique to MDD did not (OR 1.00, 95% CI 0.95–1.06;  $P$  0.97). Our replication sample was the Atherosclerosis Risk in Communities (ARIC) cohort of 7197 European-ancestry participants (1598 incident CAD cases). In ARIC, polygenic scores for MDD and loneliness were associated with hazard ratios of 1.07 (95% CI, 0.99–1.14;  $P$  = 0.07) and 1.07 (1.01–1.15;  $P$  = 0.03), respectively, and we replicated findings from the BioVU sensitivity analyses. We conclude that genetic risk factors for MDD and loneliness act pleiotropically to increase CAD risk in females.

## Introduction

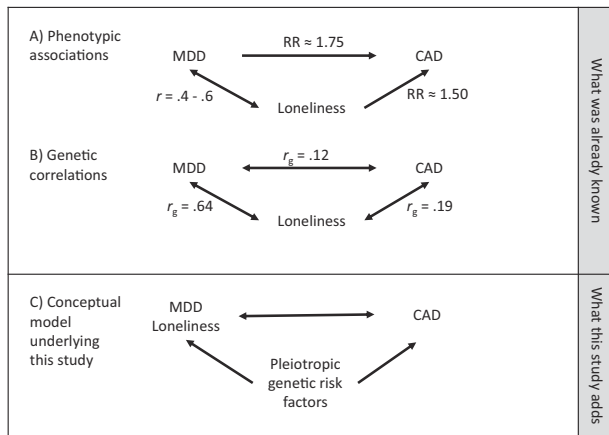
Major depressive disorder (MDD) is among the most common mental disorders, with a lifetime prevalence as high as 17% [1]. MDD and heart disease are closely intertwined. Patients with MDD have an ~75% increased risk of

cardiovascular disease [2] and 30–74% of patients who have cardiac events will also meet criteria for MDD (Fig. 1a) [3]. Understanding the complex relationship between these two traits is critical to reducing early mortality, as comorbid MDD is associated with a fivefold increased risk of cardiac mortality within 6 months of a myocardial infarction [3]. In addition to the frequently documented, but poorly understood, relationship between MDD and cardiac health, recent evidence also strongly suggests that chronic loneliness—independent from MDD—similarly increases the risk of heart disease and early mortality. Analyses of ~500,000 participants in the UK biobank showed that loneliness increased the risk of cardiovascular disease incidence by 50% [4] and

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41380-019-0614-y>) contains supplementary material, which is available to authorized users.

✉ Lea K. Davis  
lea.k.davis@vumc.org

Extended author information available on the last page of the article.



**Fig. 1** Results from large-scale genome-wide association studies (GWAS) of major depressive disorder (MDD) [11], loneliness [10], and coronary artery disease (CAD) [12] can be used to disentangle relationships between the three traits. **a** Genome-wide genetic correlations show that all three traits share some genetic risk factors. **b** Mendelian randomization experiments using the top associations from GWAS as genetic instruments found no causal effect of CAD on MDD on loneliness, while the reverse tests were likely under-powered. **c** This study used polygenic scores for MDD and loneliness in electronic health records linked to DNA samples and found pleiotropic relationships between MDD, loneliness, and CAD even after accounting for comorbidities and conventional heart disease risk factors

cardiovascular disease death by 30% [5]. This finding among others [6] has prompted public health initiatives to investigate the physical and mental impact of loneliness. Chronic loneliness reflects a discrepancy between a person's desired and perceived social connectedness, and is experienced by up to 22% of adults [7].

Loneliness and depressive symptoms are phenotypically correlated (Pearson correlation of 0.4–0.6) [8, 9], but are believed to represent distinct underlying phenomena. Like MDD, loneliness is a heritable trait [10] and the mechanisms linking loneliness to heart disease are incompletely understood.

Summary statistics from large-scale genome-wide association studies (GWAS) show that MDD [11], loneliness [10], and coronary artery disease (CAD) [12], the most common type of heart disease, are genetically correlated, suggesting shared genetic risk factors (Fig. 1b). However, genetic correlations based on GWAS summary statistics can be difficult to interpret, as each study typically ascertains cases on only a single disease without adjusting for or matching on phenotypic comorbidities. While this approach is optimal in GWAS for many reasons, it introduces the possibility of “phenotypic hitchhiking” in which a comorbid trait (e.g., MDD) is unintentionally selected during the ascertainment of the index trait (e.g., CAD). Thus, two heritable phenotypes (e.g., MDD and CAD) that share common environmental risk factors (e.g., obesity) but no genetic risk factors can appear genetically correlated in an analysis based solely on GWAS summary statistics.

Dense, longitudinal clinical data from large genotyped samples provide new opportunities to disentangle shared genetic risk factors from phenotypic hitchhiking. We hypothesized that pleiotropic genetic risk factors in part explained the phenotypic associations and genetic correlations between MDD, loneliness, and CAD (Fig. 1c). We assessed this hypothesis by testing polygenic scores for MDD, loneliness, and CAD for association with clinical diagnoses of MDD and CAD, before and after adjusting for known disease risk factors and comorbidity patterns, in a large collection of patient electronic health records (EHR) linked to genome-wide genotyping. Polygenic scores aggregate the effects of many genetic variants into a single measure of genetic liability to a trait, and can be calculated for all individuals with genome-wide genetic data if their ancestry is similar to the ancestry of the participants in the external GWAS from which the individual genetic variant effect estimates are derived.

This study included five aims. The first was to identify the medical morbidity patterns associated with polygenic scores for MDD and loneliness in the Vanderbilt University Medical Center (VUMC) EHR and accompanying biobank, BioVU. Second, we undertook a targeted analysis of CAD, and determined whether genetic risk for MDD and loneliness remained associated with CAD in the absence of a clinical diagnosis of MDD. Third, we examined whether the risk of CAD conferred by MDD and loneliness risk scores could be attenuated by accounting for multiple conventional CAD risk factors. Fourth, given the well-established sex differences in MDD incidence [1], and in CAD mortality, presentation, and risk factors [13–15], we repeated all analyses in males and females separately. Fifth, we replicated our findings in a large prospectively collected population cohort, the Atherosclerosis Risk in Communities (ARIC) study. Our results provided strong evidence that pleiotropic risk factors for MDD and loneliness increase the risk of CAD, with more pronounced effects in females than in males.

## Methods

### Study populations

The VUMC biobank, BioVU, served as the discovery population. BioVU includes more than 285,000 inpatients and outpatients seen at VUMC whose DNA was extracted from whole blood, and linked to de-identified EHR data spanning 1990–2017 [16]. The BioVU Consent form is provided to patients in the clinic environments at VUMC. The form states policies on data sharing and privacy, and should a signature be obtained, makes any blood leftover from clinical care eligible for BioVU banking. The VUMC

Institutional Review Board oversees BioVU and approved this project. The ARIC cohort served as our replication population [17]. The study included 15,792 participants aged 45–64 years in 1987 recruited from four geographic areas in the United States, assessed at baseline for heart disease risk factors, and followed until December 31, 2004 for health outcomes [18]. Use of de-identified ARIC data was approved by the VUMC Institutional Review Board. De-identified ARIC phenotype and genotype data were obtained from dbGaP and analyzed per requirements laid out in the dbGaP Data Use Agreement.

### Genotyping and quality control

A subset of BioVU patients ( $n = 24,262$ ) was genotyped as part of various institutional and investigator-initiated projects on the Illumina MEGA<sup>EX</sup> platform, which contains more than two million markers. Quality control proceeded as previously described [19]. Genotypes were imputed using SHAPEIT [20]/IMPUTE4 [21] with the 1000 genomes phase I reference panel, and variants with  $\text{INFO} < 0.3$  were excluded. A subset of SNPs in linkage disequilibrium was used to calculate relatedness and principal components of ancestry using multidimensional scaling in PLINK v1.9 [22]. We randomly removed one individual from pairs of highly related individuals ( $\text{pihat} > 0.1$ ) to avoid spurious results driven by cryptic relatedness, and restricted to a homogenous population of European descent defined by principal components of ancestry to avoid population stratification effects, leaving 18,385 individuals for analyses. Samples were genotyped in five batches, and variants were removed if allele frequencies differed significantly ( $P < 5 \times 10^{-5}$ ) between any batch and the rest of the sample. Finally, we filtered multiallelic and structural variants, converted dosage data to hard genotype calls, and excluded variants with certainty  $< 0.9$  or  $\text{INFO} < 0.95$ , resulting in 5,218,407 high quality SNPs across the autosomes.

Genetic data from 13,113 ARIC participants were downloaded from dbGaP (phs000280.v3.p1). Genotypes were measured on the Affymetrix 6.0 SNP array, and data were cleaned as previously described [23]. Genetic ancestry was assigned using STRUCTURE [24] in conjunction with HapMap reference populations, and European ancestry was defined as  $> 90\%$  probability of being in the CEU cluster. Genotypes were imputed to the October 2014 Phase 3 release of the 1000 Genomes cosmopolitan reference haplotypes [25] using SHAPEIT [20]/IMPUTE2 [21] and we excluded SNPs with  $\text{INFO} < 0.7$ , multiallelic SNPs, and structural variants. Dosage data were converted to hard genotyping calls, excluding variants with certainty  $< 0.9$ , leaving 6,569,625 high quality SNPs for analysis. Inter-individual relatedness and principal components of ancestry were calculated by GCTA [26] and we randomly

excluded one individual from pairs of related individuals ( $\text{pihat} > 0.05$ ), leaving 6975 unrelated European ancestry participants.

### Phenotype data

To facilitate the investigation of comorbidity patterns across the medical phenome, we mapped International Classification of Diseases, ninth edition (ICD-9) billing codes in the EHR to phecodes, which are the higher-order representations of disease categories. ICD-9 codes were mapped to 1814 phecode categories according to the Phecode Map v1.2 (<https://phewascatalog.org/phecodes>), as implemented in the PheWAS R package v0.12 [27]. Patients were assigned to the case group for a given phecode if they had at least two different ICD-9 codes that mapped to a given phecode, or if they had at least two separate occurrences (i.e., on different days) of a single ICD-9 code that mapped to the given phecode, both of which are validated strategies to improve the positive predictive value of phecodes [28]. The control group excluded patients with only one component ICD-9 code, or with one or more ICD-9 codes that mapped to related phecodes (as defined by the Phecode Map v1.2).

Heart disease and its most common manifestation, CAD, were a focus of our analysis. However, no single phecode was diagnostic of CAD. Therefore, we defined CAD in BioVU patients by a random forest (machine learning) classifier [29] that integrated data from across the EHR (Supplementary Methods; Supplementary Table 1; Supplementary Figs. 1–3). Age in cases was defined by the age of first CAD-defining feature in the EHR, while in controls, age was defined by the age at last encounter. Data on CAD risk factors were also extracted from the EHR either from structured data or via text-mining algorithms (Supplementary Methods). CAD risk factors included body mass index (BMI), hypertension, type 2 diabetes diagnosis, pre-medication blood levels of high- and low-density lipoprotein cholesterol (HDL and LDL) and of triglycerides [30] (lipid-altering medications are listed in Supplementary Table 2), smoking history, and socio-economic status.

A key research question was whether associations with polygenic scores for MDD and loneliness persisted in people without a clinical MDD diagnosis. We defined MDD by the phecode 296.22 “Major depressive disorder”, which included both single episode and recurrent major depressive episodes that are severe, moderate, or of unspecified severity. We also defined broad categories of mental illness, to adjust for the possibility that any degree of mental illness could increase CAD risk simply due to lifestyle factors that may be correlated with having the mental illness itself. We defined a milder depressive symptoms phenotype by one or more depression ICD-9 codes (296.2 and 296.3, including

all fifth digit subclassifications, and 311), and as a final test of the robustness of associations, we excluded patients with any psychiatric symptoms (Supplementary Methods). Due to limitations of collecting medication data (e.g., incomplete records, noncompliance, etc.) we did not extract data on psychiatric medications but did model their downstream consequences (i.e., increased BMI).

Phenotype data for ARIC samples came from measurements aggregated by the GENEVA substudy (pht000114.v2.p1), downloaded with permission from dbGaP [23]. Eligible participants were those with data on incident CAD, defined as myocardial infarction, fatal coronary heart disease, silent myocardial infarction detected by electrocardiography, or revascularization procedure. We additionally extracted data on sex, age at first visit, and time to event (CAD or censoring in controls), as well as first visit data on BMI, waist girth, smoking status, hypertension medication use, systolic and diastolic blood pressure, type 2 diabetes diagnosis, highest level of education, use of cholesterol-lowering medication and other medications that secondarily affect cholesterol, and blood measurements of HDL, LDL, and triglycerides (Supplementary Table 3). We defined hypertension by the variable hypertension medication use, or systolic blood pressure >130 mmHg, or diastolic blood pressure >80 mmHg [31]. MDD was not assessed in ARIC participants because ARIC was designed to study cardiovascular outcomes. Therefore, we were only able to test the MDD and loneliness polygenic scores for association with CAD but were not able to adjust those analyses for the presence of clinical MDD.

## Statistical analyses

Polygenic scores for MDD and loneliness were computed using the PRSice software package [32] (Supplementary Methods) for each individual in BioVU and ARIC as follows:

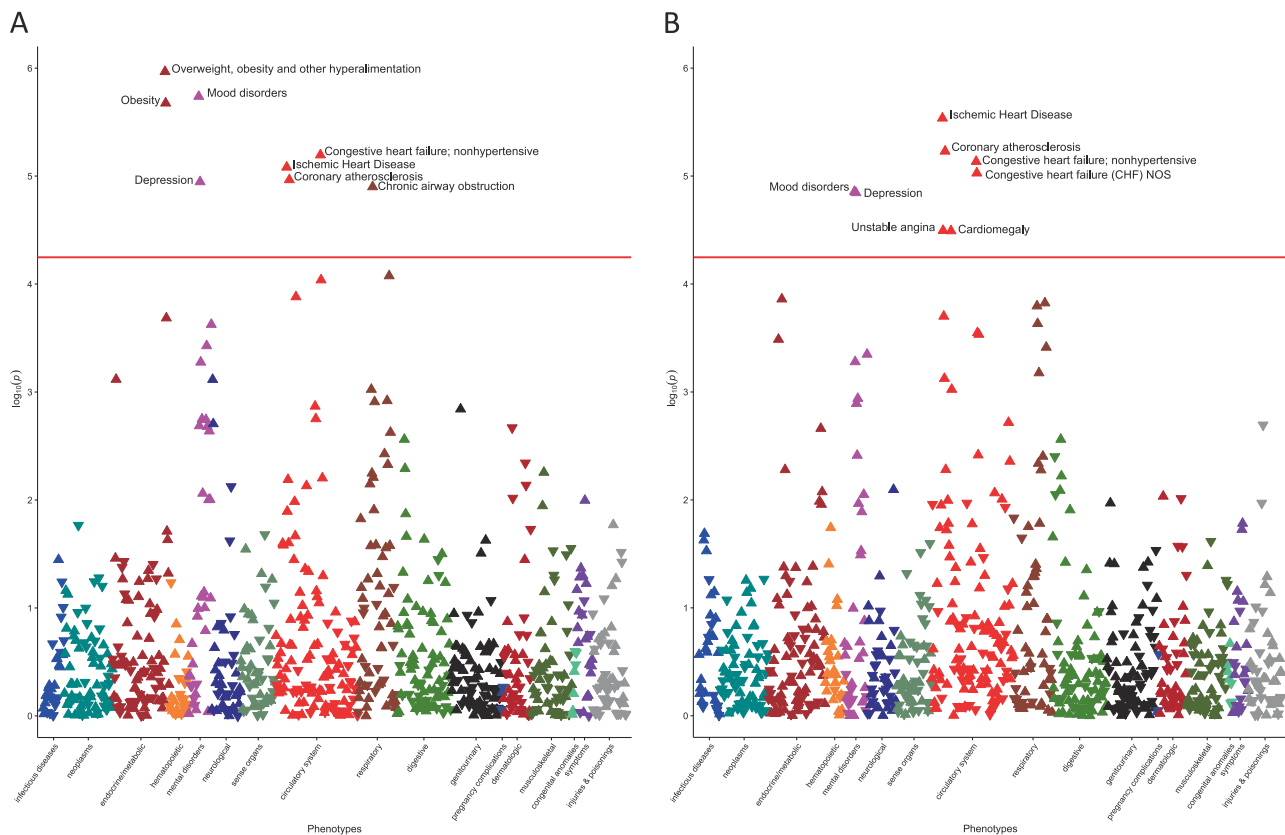
$$\text{Polygenic score} = \sum_{i=1}^{\#SNPs} (w_i \times [\text{SNP genotype}]_i)$$

where the SNP genotype was coded as 0, 1 or 2 and  $w_i$  was the  $\beta$  coefficient (representing effect size of the marginal association between the SNP and the phenotype) from the Psychiatric Genomics Consortium meta-GWAS of MDD [11], or the Lonely Consortium meta-GWAS of loneliness [10]. We selected SNPs into the polygenic score if their association  $P$  in the meta-GWAS was below a specified threshold. In the phenome-wide association study, we selected an  $P$  threshold of 0.05, a value previously found to generate scores that maximized the prediction of MDD [11] and other psychiatric traits [33], and included 392,372 and 379,906 SNPs in the polygenic scores of MDD and

loneliness, respectively. In targeted analyses of CAD, we selected the  $P$  threshold at which the polygenic score was most strongly associated with CAD risk (i.e., at an  $P$  threshold different than 0.05). This approach improved our power to detect associations with CAD by maximizing the number of pleiotropic SNPs included in the MDD and loneliness polygenic scores. We determined  $P$  thresholds of  $9.0005 \times 10^{-4}$  for MDD (n SNPs = 1387) and  $2.50005 \times 10^{-3}$  for loneliness (n SNPs = 2253) by iterating over  $P$  from  $5 \times 10^{-8}$  to 1 in increments of  $5 \times 10^{-5}$  and assessing fit via Nagelkerke's pseudo  $R^2$  (Supplementary Fig. 4). Polygenic scores calculated at these best fit  $P$  thresholds were used in all sensitivity and replication analyses of CAD to minimize false findings due to overfitting. To adjust for the genetic correlation between MDD and loneliness (Fig. 1b) [10], we used the multitrait-based conditional and joint analysis (mtCOJO) package [34], which uses only meta-GWAS summary statistics to statistically isolate independent genetic associations. Using mtCOJO, we calculated SNP associations with MDD conditional on loneliness (MDD|loneliness), and vice-versa (loneliness|MDD), and calculated polygenic scores from the MDD|loneliness and loneliness|MDD summary statistics using the best fit  $P$  thresholds above. All polygenic scores were generated by the default pruning and thresholding parameters in PRSice v2 [32] (Supplementary Methods) and were scaled to have an SD of 1. Risk ratios (odds or hazard) reflect the risk increase per 1-SD increase in the polygenic score.

Phenome-wide association studies were conducted with the PheWAS R package v0.12 [27]. We required phecodes to have at least 100 cases, and included covariates for sex, median age across the EHR, genotyping batch, and the first 10 principal components of ancestry.

Targeted analyses of CAD in BioVU employed logistic regression. Minimally adjusted models included covariates for sex, age, genotype batch, and the first ten principal components of ancestry. Fully adjusted models additionally included covariates for BMI, hypertension, smoking, type 2 diabetes, blood measurements of HDL, LDL, and triglycerides, highest level of education, a polygenic score for CAD (Supplementary Methods), and depressive symptoms. In ARIC, we modeled incident CAD using a Cox proportional hazards model and included sex, baseline age, and the first ten principal components of ancestry in the minimally adjusted models. Fully adjusted models included additional covariates for waist girth, smoking status, hypertension medication use, systolic blood pressure, type 2 diabetes, highest level of education, use of cholesterol-lowering medication and other medications that secondarily affect cholesterol, and blood measurements of HDL, LDL, and triglycerides. Although these covariates did not exactly match those we controlled for in BioVU, (i.e., waist girth



**Fig. 2** Results from phenome-wide association studies of polygenic scores for MDD (**a**) and loneliness (**b**) in BioVU. The red line denotes the Bonferroni threshold for statistical significance ( $0.05/882 = 5.67 \times 10^{-5}$ ), and phenome-wide significant phecodes are labeled. Upward

triangles indicate increased odds for a given phecode per 1-SD increased risk in the polygenic score, while downward triangles indicate reduced odds of a given phecode. Interactive plots can be viewed at: [https://sealockj.shinyapps.io/mdd\\_loneliness\\_cad\\_interactive/](https://sealockj.shinyapps.io/mdd_loneliness_cad_interactive/)

instead of BMI, hypertension medication use and systolic blood pressure instead of hypertension diagnosis), we used them preferentially since they are better measures of the underlying risk factors, but were rarely reported (waist girth) or error-prone (systolic blood pressure) in the EHR. In the fully adjusted models for both BioVU and ARIC, we tested for interaction between sex and the polygenic score for MDD or loneliness and report interaction  $P < 0.05$ . In addition, we provide the results of sex-stratified analyses, as effect estimates may differ even in the absence of a statistically significant interaction.

All analyses were completed in R version 3.4.3 or 3.5, and code is available upon request.

## Results

Our analyses included 18,385 BioVU patients. These patients had a median EHR length of 9.91 years (range <1–27 years; Supplementary Table 4), the average median age across the EHR was 57.2 years (SD 17.6), and 50.9% of patients were female. Nearly 5% of patients were diagnosed with MDD, 24.9% had depressive symptoms, and 39.5%

had evidence of any psychiatric symptoms. While the genome-wide genetic correlation between MDD and loneliness was 0.64 [10], the polygenic scores for MDD and loneliness had a Pearson correlation of 0.18 (Supplementary Table 5).

Our first aim was to identify medical morbidities associated with these polygenic scores using a phenome-wide association study. We analyzed 882 phecodes with at least 100 cases, and used a Bonferroni-corrected phenome-wide significance threshold of  $0.05/882 = 5.67 \times 10^{-5}$ . This threshold, however, is likely over-conservative because it incorrectly assumes independence between phecodes. The phenome-wide association study of the polygenic score for MDD identified phenome-wide significant associations with “Mood disorders” (odds ratio [OR], 1.10 [95% CI, 1.06–1.15];  $P = 1.83 \times 10^{-6}$ ), and “Depression” (OR, 1.10 [95% CI, 1.05–1.15];  $P = 1.13 \times 10^{-5}$ ), parent codes for MDD, as expected (Fig. 2a; interactive plots available at: [https://sealockj.shinyapps.io/mdd\\_loneliness\\_cad\\_interactive/](https://sealockj.shinyapps.io/mdd_loneliness_cad_interactive/)). The remaining six phecodes that were significant after Bonferroni correction were related to acute and chronic heart diseases and their associated risk factors, including “Ischemic heart disease” (OR, 1.09 [95% CI, 1.05–1.14];

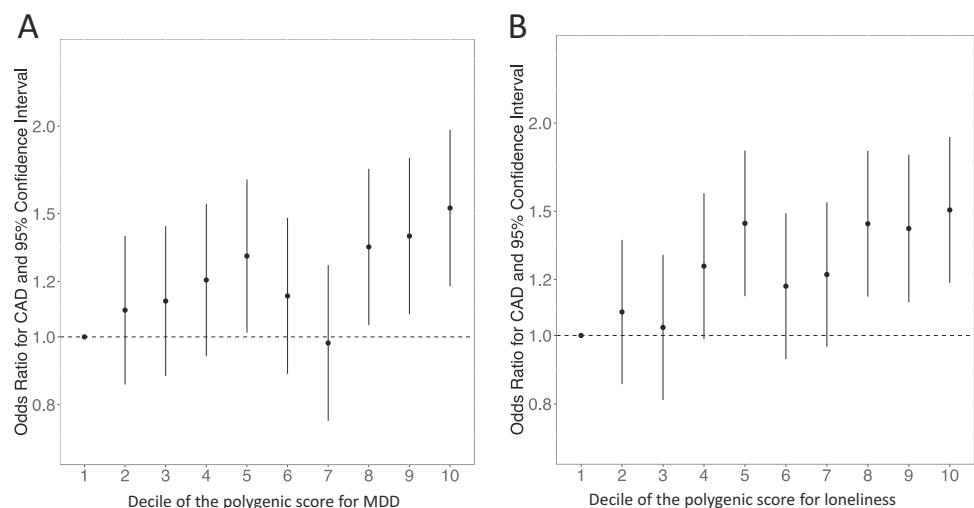
$P = 8.28 \times 10^{-6}$ ) and “Coronary atherosclerosis” (OR, 1.10 [95% CI, 1.05–1.14];  $P = 1.08 \times 10^{-5}$ ). The phenome-wide association study of the polygenic score for loneliness (Fig. 2b) also found significant associations with “Mood disorders” (OR, 1.09 [95% CI, 1.05–1.14];  $P = 1.39 \times 10^{-5}$ ) and “Depression” (OR, 1.10 [95% CI, 1.05–1.14];  $P = 1.43 \times 10^{-5}$ ). The top associations, however, were with heart disease phenotypes, specifically, “Ischemic heart disease” (OR, 1.10 [95% CI, 1.06–1.14];  $P = 2.91 \times 10^{-6}$ ) and “Coronary atherosclerosis” (OR, 1.10 [95% CI, 1.06–1.15];  $P = 5.89 \times 10^{-6}$ ). In contrast, a polygenic score for CAD (Supplementary Methods) was not associated with MDD (OR, 0.93 [95% CI, 0.87–1.00];  $P = 0.07$ ), or other mental disorder diagnoses ( $P > 0.05$ ; see above link to interactive plots and Supplementary Fig. 5), replicating previously published findings [35].

Motivated by these results, for our second aim, we targeted CAD, to better understand the contribution of genetic risk factors for MDD and loneliness to the shared comorbidity patterns implicating CAD. Our machine learning algorithm identified 3893 cases and 4197 controls in BioVU (Supplementary Table 6). In minimally adjusted models, each SD increase in the polygenic scores for MDD and loneliness, respectively, increased the odds of CAD by 1.11 (95% CI, 1.04–1.18;  $P = 8.43 \times 10^{-4}$ ) and 1.13 (95% CI, 1.08–1.20;  $P = 4.51 \times 10^{-6}$ ). These findings mirrored those from the phenome-wide association study and indicated that our results were not biased by the classification ability of our algorithm. Stratified by deciles, patients with polygenic scores for MDD and loneliness in the top versus bottom deciles had, respectively, a 1.53-fold (95% CI, 1.18–1.98;  $P = 1.2 \times 10^{-3}$ ) and 1.51-fold (95% CI, 1.19–1.91;  $P = 7.4 \times 10^{-4}$ ) greater risk of CAD (Fig. 3). These associations persisted even after excluding patients with a clinical diagnosis of MDD, depressive symptoms, or any psychiatric symptoms (Table 1).

For aims 3 and 4, we tested these associations in multivariable models overall, and stratified by sex, and found that associations were robust to adjustment for conventional risk factors in females, but not in males (Fig. 4a, b; polygenic score for loneliness by sex interaction  $P < 0.05$ ). We next accounted for the genetic correlation between MDD and loneliness using the mtCOJO package [34] (Pearson correlations between polygenic scores are in Supplementary Table 5). The polygenic score for MDD|loneliness was not associated with CAD risk (Fig. 4c), whereas the score for loneliness|MDD remained associated with CAD risk in females, even in the fully adjusted model (Fig. 4d). The polygenic score for loneliness|MDD was associated with a 1.23-fold (95% CI, 1.02–1.47;  $P = 0.017$ ) increased odds of CAD in females in the fully adjusted model that maximized patient inclusion by including categories for missing values. The association was equivalent, although less statistically significant (OR, 1.23 [95% CI, 0.94–1.60];  $P = 0.126$ ) in the fully adjusted model that excluded the 699 female BioVU patients with missing covariate data.

Our fifth aim was to replicate our findings in an independent population. The ARIC sample included 7197 participants of European ancestry with clean genotyping data, followed for a median of 16.0 years (range 0.05–18.1), over which time 923 incident CAD cases were recorded (Supplementary Table 7). The polygenic score for MDD was associated with a hazard ratio (HR) of 1.07 (95% CI, 0.99–1.14;  $P = 0.07$ ), as was the polygenic score for loneliness (HR, 1.07 [95% CI, 1.01–1.15];  $P = 0.03$ ). Associations were robust to confounder adjustment in females but not in males (Supplementary Fig. 6), and the polygenic score for loneliness|MDD was associated with a CAD HR of 1.14 (95% CI, 1.01–1.29;  $P = 0.02$ ) in females versus 0.98 (95% CI, 1.02–1.29;  $P = 0.60$ ) in males in the fully adjusted model (interaction  $P < 0.05$ ).

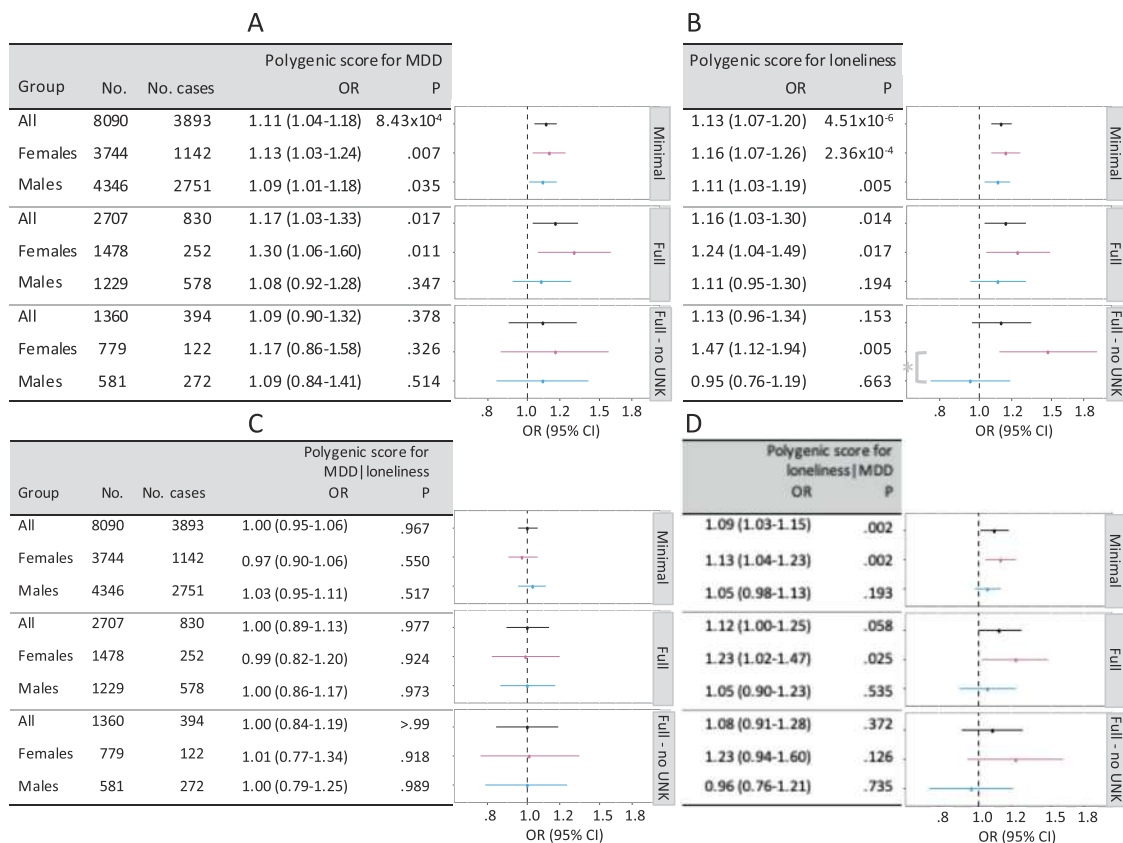
**Fig. 3** Odds of CAD by decile of the polygenic score for MDD (a) and loneliness (b) in BioVU. The referent group in all calculations is the lowest polygenic score decile



**Table 1** Associations between polygenic scores for major depressive disorder (MDD) and loneliness with coronary artery disease (CAD) after excluding patients with a diagnosis of MDD, depressive symptoms, or any psychiatric symptoms

| Polygenic score | Patient exclusions       | N Cases with CAD | N Controls without CAD | OR (95% CI)      | P                     |
|-----------------|--------------------------|------------------|------------------------|------------------|-----------------------|
| MDD             | None                     | 3893             | 4197                   | 1.11 (1.04–1.18) | $8.43 \times 10^{-4}$ |
| MDD             | MDD                      | 3744             | 4037                   | 1.11 (1.05–1.18) | $6.54 \times 10^{-4}$ |
| MDD             | Depressive symptoms      | 2935             | 3147                   | 1.10 (1.03–1.18) | $5.76 \times 10^{-3}$ |
| MDD             | Any psychiatric symptoms | 2440             | 2551                   | 1.11 (1.03–1.19) | $9.27 \times 10^{-3}$ |
| Loneliness      | None                     | 3893             | 4197                   | 1.13 (1.08–1.20) | $4.51 \times 10^{-6}$ |
| Loneliness      | MDD                      | 3744             | 4037                   | 1.13 (1.07–1.20) | $9.22 \times 10^{-6}$ |
| Loneliness      | Depressive symptoms      | 2935             | 3147                   | 1.14 (1.08–1.22) | $2.08 \times 10^{-5}$ |
| Loneliness      | Any psychiatric symptoms | 2440             | 2551                   | 1.12 (1.05–1.20) | $8.69 \times 10^{-4}$ |

MDD major depressive disorder



**Fig. 4** Risk of CAD predicted by polygenic scores for MDD (a), loneliness (b), MDD|loneliness (c), and loneliness|MDD (d) before and after adjustment for conventional heart disease risk factors. Minimally adjusted (“Minimal”) models included sex (except sex-stratified results), age, the first ten principal components of ancestry, and genotype batch. Fully adjusted (“Full”) models included additional covariates for BMI, hypertension, smoking, type 2 diabetes, blood measurements of HDL, LDL, and triglycerides, highest level of

education, a polygenic score for CAD, and depressive symptoms. Values for smoking and highest level of education were unknown for ~50% of patients. These unknown values were retained in the “Full” model by modeling an explicit category for “missing” values, and were removed from the “Full - no UNK” model. Patients were excluded from the fully adjusted models if there were missing values for any of the other included covariates. The asterisk in b denotes sex  $\times$  polygenic score interaction  $P < .05$

## Discussion

This study investigated the phenome-wide consequences of a genetic liability to MDD and loneliness in BioVU, which

is a deeply phenotyped large-scale EHR collection with linked genotype data. We discovered strong associations between polygenic scores for MDD and loneliness with heart disease, and our targeted analysis of CAD revealed

that patients with polygenic scores in the highest decile for either MDD or loneliness had a more than 50% greater risk of CAD compared with patients with scores in the lowest decile. Conversely, a polygenic score for CAD was not associated with MDD or other psychiatric phenotypes, in accordance with a previous study of the UK biobank population [35]. Epidemiological associations between MDD and loneliness and heart disease are known [36]; here we showed that the association is partly attributable to genetics. Similar results were observed in ARIC participants, who were prospectively surveyed for heart disease outcomes using traditional epidemiological approaches.

An unique feature of our study was the extensive sensitivity analyses that provided more insight into the reported genetic and epidemiological correlations between MDD, loneliness, and CAD. Genetic correlations alone can be difficult to interpret because they may be influenced by phenotypic hitchhiking. For example, if MDD is comorbid with heart disease, then a GWAS of MDD based on prevalent cases will also ascertain patients with heart disease, which could then induce a genetic correlation between MDD and an independent GWAS of heart disease. Any GWAS that ascertains prevalent cases will be similarly enriched for comorbidities. In the absence of large GWAS meta-analyses of incident cases, we tackled this problem by controlling for multiple clinical risk factors, known CAD risk variants, and genetic correlations, resulting in strong evidence that genes conferring risk for MDD and loneliness exert pleiotropic effects on CAD.

MDD and loneliness are genetically correlated [10]. Our conditional analyses, however, suggested that genetic risk factors specific to loneliness increased CAD risk independent of the genetic risk factors that are shared between MDD and loneliness (Fig. 1). This finding is congruent with the physiological effects of chronic loneliness in humans and in animal models [37]. Loneliness induces a state of self-preservation in anticipation of being without the protection of others: it triggers depressive symptoms that signal the need for support and connection from peers [9]; disrupts sleep to maintain a state of alertness at night [38]; raises blood pressure [39]; and activates the hypothalamic pituitary adrenal axis [37, 40–42], which regulates cortisol, a key hormone in stress reactions, metabolism, digestion, immunity, and energy storage. While these responses may be advantageous for being alone in the short term, health problems ensue when loneliness is chronic and the biological response is sustained [37].

Both MDD and heart disease exhibit significant sex differences in prevalence and presentation [1, 13, 15], and heart disease remains the number one killer of females in the United States [14]. Our analyses revealed that genetic risk for MDD and loneliness conferred a higher risk of CAD

in females than in males, and that the risk in females was robust to adjustment for the major known risk factors. There are two possible, non-mutually exclusive, explanations for this finding. First, genetic predisposition to MDD or loneliness may be a chronic risk factor for females but an acute risk factor for males, as was previously suggested in a study of 30,000 twins from the Swedish population-based twin registry [43]. Second, the conventional risk factors identified in most epidemiological studies may be poorer predictors of CAD for females than they are for males. Non-traditional and female-specific CAD risk factors (e.g., rheumatoid arthritis, preterm delivery) may contribute to sex differences in CAD prevalence, presentation, and mortality [15], but have only recently been interrogated and so we did not include them in our multivariable models. Future work is warranted to investigate the role of pleiotropic genes on both heart disease and MDD, especially in females where the effects are strongest, and the etiology of heart disease more obscure.

## Limitations

EHR data reflect real clinical use patterns, but some data can be missing or inaccurate. We addressed this potential limitation by implementing a carefully curated CAD phenotype, and conducted sensitivity analyses excluding people with unknown covariate values. In addition, the VUMC inpatient and outpatient population included in our study may be sicker than the general population. Nonetheless, we replicated our findings in a prospective epidemiological cohort ascertained from the general population. Even though MDD was not assessed in ARIC participants, which hampered our ability to replicate results from aim 2, we replicated all other aims and findings, including that polygenic scores for MDD and loneliness were associated with CAD even after adjusting for conventional heart disease risk factors. Another limitation of our study was its restriction to individuals of European ancestry, but this decision was necessitated by the ancestry of patients in the meta-GWAS used to build the polygenic scores. The relevance of our findings to individuals of other ancestries, therefore, is unknown [44, 45].

## Conclusions

Mental health has historically been siloed away from the rest of medicine, resulting in a poor understanding of the relationship between mental and physical health. This EHR-based study showed that genetic risk for MDD and loneliness were just as strongly associated with heart disease as with depression itself. Moreover, the increased risk of CAD persisted in females after adjusting for psychiatric



symptoms and multiple other risk factors, suggesting that the excess risk is due to pleiotropic genetic effects and is not simply a behavioral consequence of depression. Identification of these pleiotropic genes could advance both precision medicine and drug development.

**Acknowledgements** JD is supported by the Canadian Institutes of Health Research (award MFE-142936). RTL is supported by AHA grant 17SFRN33520017 and NIH 1R01HL140074-01. MFL is supported by PO1HL116263 and R01HL127173. SS-R is supported by the Frontiers of Innovation Scholars Program (FISP; #3-P3029), the Interdisciplinary Research Fellowship in NeuroAIDS (IRFN; MH081482), and a pilot award from DA037844. SS-R and AAP are supported by funds from the California Tobacco-Related Disease Research Program (TRDRP; Grant nos. 28IR-0070 and T29KT0526). AAP is supported by P50DA037844. AA and DIB acknowledge NIH R37 AG033590-08 grant to J Cacioppo. Support for NJC was provided by R01MH113362, U54MD010722, and U01HG009086. GC is supported by NIH UL1TR000427. JDM is supported by AHA grant 16FTF30130005. LKD is supported by NIH 1R01MH118233-01 and 1R56MH120736-01. This project was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN. The datasets used for this project were obtained from Vanderbilt University Medical Center's Synthetic Derivative, which is supported by numerous sources: institutional funding, private agencies, and federal grants. The project described was supported by the National Center for Research Resources, Grant UL1 RR024975-01, and is now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Data on coronary artery disease/myocardial infarction have been contributed by CARDIOGRAMplusC4D investigators and have been downloaded from [www.CARDIOGRAMPLUSC4D.ORG](http://www.CARDIOGRAMPLUSC4D.ORG). The authors thank the staff and participants of the ARIC study for their contributions. ARIC is supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). ARIC/GENEVA was supported by NHGRI grant U01HG004402 (E. Boerwinkle).

## Compliance with ethical standards

**Conflict of interest** PF and SLE are employees of 23andMe, Inc. No other disclosures are reported.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. *Annu Rev Public Health*. 2013;34:119–138.
2. Correll CU, Solmi M, Veronese N, Bortolato B, Rosson S, Santonastaso P, et al. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. *World Psychiatry*. 2017;16:163–180.
3. Dhar AK, Barton DA. Depression and the link with cardiovascular disease. *Front Psychiatry*. 2016;7:33.
4. Hakulinen C, Pulkki-Raback L, Virtanen M, Jokela M, Kivimaki M, Elovainio M. Social isolation and loneliness as risk factors for myocardial infarction, stroke and mortality: UK Biobank cohort study of 479 054 men and women. *Heart*. 2018;104:1536–42.
5. Elovainio M, Hakulinen C, Pulkki-Raback L, Virtanen M, Josefsson K, Jokela M, et al. Contribution of risk factors to excess mortality in isolated and lonely individuals: an analysis of data from the UK Biobank cohort study. *Lancet Public Health*. 2017;2:e260–e266.
6. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. *Perspect Psychol Sci*. 2015;10:227–37.
7. Qualter P, Vanhalst J, Harris R, Van Roekel E, Lodder G, Bangee M, et al. Loneliness across the life span. *Perspect Psychol Sci*. 2015;10:250–64.
8. Matthews T, Danese A, Wertz J, Odgers CL, Ambler A, Moffitt TE, et al. Social isolation, loneliness and depression in young adulthood: a behavioural genetic analysis. *Soc Psychiatry Psychiatr Epidemiol*. 2016;51:339–48.
9. Cacioppo JT, Hawkley LC, Thisted RA. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago Health, Aging, and Social Relations Study. *Psychol Aging*. 2010;25:453–63.
10. Abdellaoui A, Sanchez-Roige S, Sealock J, Treur J, Dennis J, Fontanillas P et al. Phenome-wide investigation of health outcomes associated with genetic predisposition to loneliness. *bioRxiv*. 2018. <https://doi.org/10.1101/468835>.
11. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50:668–81.
12. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47:1121–30.
13. Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, et al. Heart disease and stroke statistics-2012 update: a report from the American Heart Association. *Circulation*. 2012;125:e2–e220.
14. Xu J, Murphy SL, Kochanek KD, Bastian BA. Deaths: final data for 2013. *Natl Vital Stat Rep*. 2016;64:1–119.
15. Garcia M, Mulvagh SL, Merz CN, Buring JE, Manson JE. Cardiovascular disease in women: clinical perspectives. *Circ Res*. 2016;118:1273–93.
16. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balsler JR, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharm Ther*. 2008;84:362–9.
17. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989;129:687–702.
18. Knopman DS, Gottesman RF, Sharrett AR, Wruck LM, Windham BG, Coker L, et al. Mild cognitive impairment and dementia

- prevalence: the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS). *Alzheimers Dement.* 2016;2:1–11.
19. Ruderfer DM, Walsh CG, Aguirre MW, Tanigawa Y, Ribeiro JD, Franklin JC et al. Significant shared heritability underlies suicide attempt and clinically predicted probability of attempting suicide. *Mol Psychiatry.* 2019. [Epub ahead of print].
  20. Delaneau O, Marchini J, Genomes Project C, Genomes Project C. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun.* 2014;5:3934.
  21. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5:e1000529.
  22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
  23. Mosley JD, Shoemaker MB, Wells QS, Darbar D, Shaffer CM, Edwards TL et al. Investigating the genetic architecture of the PR interval using clinical phenotypes. *Circ Cardiovasc Genet.* 2017;10:e001482.
  24. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155:945–59.
  25. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* 2015;526:68–74.
  26. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88:76–82.
  27. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics.* 2014;30:2375–6.
  28. Wei WQ, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. *J Am Med Inf Assoc.* 2016;23:e20–27.
  29. Breiman L. Random forests. *Mach Learn.* 2001;45:5–32.
  30. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. *J Am Med Inf Assoc.* 2010;17:19–24.
  31. Whelton PK, Carey RM, Aronow WS, Casey DE Jr., Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APHA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American college of cardiology/American Heart Association task force on clinical practice guidelines. *J Am Coll Cardiol.* 2018;71:e127–e248.
  32. Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. *Bioinformatics.* 2015;31:1466–8.
  33. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014;511:421–7.
  34. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun.* 2018;9:224.
  35. Khandaker GM, Zuber V, Rees JMB, Carvalho L, Mason AM, Foley CN et al. Shared mechanisms between coronary heart disease and depression: findings from a large UK general population-based cohort. *Mol Psychiatry.* 2019. [Epub ahead of print].
  36. Wyman L, Crum RM, Celentano D. Depressed mood and cause-specific mortality: a 40-year general community assessment. *Ann Epidemiol.* 2012;22:638–43.
  37. Cacioppo JT, Cacioppo S, Cole SW, Capitano JP, Goossens L, Boomsma DI. Loneliness across phylogeny and a call for comparative studies and animal models. *Perspect Psychol Sci.* 2015;10:202–12.
  38. Cacioppo JT, Hawkley LC, Berntson GG, Ernst JM, Gibbs AC, Stickgold R, et al. Do lonely days invade the nights? Potential social modulation of sleep efficiency. *Psychol Sci.* 2002;13:384–7.
  39. Hawkley LC, Thisted RA, Masi CM, Cacioppo JT. Loneliness predicts increased blood pressure: 5-year cross-lagged analyses in middle-aged and older adults. *Psychol Aging.* 2010;25:132–41.
  40. Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, et al. No health without mental health. *Lancet.* 2007;370:859–77.
  41. Cacioppo S, Grippo AJ, London S, Goossens L, Cacioppo JT. Loneliness: clinical import and interventions. *Perspect Psychol Sci.* 2015;10:238–49.
  42. Brown EG, Gallagher S, Creaven AM. Loneliness and acute stress reactivity: a systematic review of psychophysiological studies. *Psychophysiology.* 2018;55:e13031.
  43. Kendler KS, Gardner CO, Fiske A, Gatz M. Major depression and coronary artery disease in the Swedish twin registry: phenotypic, genetic, and environmental sources of comorbidity. *Arch Gen Psychiatry.* 2009;66:857–63.
  44. Duncan L, Hanyang S, Gelaye B, Ressler K, Feldman M, Peterson R et al. Analysis of polygenic score usage and performance across diverse human populations. *bioRxiv.* 2018. <https://doi.org/10.1101/398396>.
  45. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale B, Daly M. Hidden 'risk' in polygenic scores: clinical use today could exacerbate health disparities. *bioRxiv* 2018. <https://doi.org/10.1101/441261>.

## Affiliations

Jessica Dennis<sup>1,2</sup> · Julia Sealock<sup>1,2</sup> · Rebecca T. Levinson<sup>2,3</sup> · Eric Farber-Eger<sup>4</sup> · Jacob Franco<sup>5</sup> · Sarah Fong<sup>1,2</sup> · Peter Straub<sup>1,2</sup> · Donald Hucks<sup>1,2</sup> · Wen-Liang Song<sup>3</sup> · MacRae F. Linton<sup>3</sup> · Pierre Fontanillas<sup>6</sup> · Sarah L. Elson<sup>6</sup> · Douglas Ruderfer<sup>1,2</sup> · Abdel Abdellaoui<sup>7,8</sup> · Sandra Sanchez-Roige<sup>9</sup> · Abraham A. Palmer<sup>9,10</sup> · Dorret I. Boomsma<sup>11,12</sup> · Nancy J. Cox<sup>1,2</sup> · Guanhua Chen<sup>13</sup> · Jonathan D. Mosley<sup>14</sup> · Quinn S. Wells<sup>3</sup> · Lea K. Davis<sup>1,2</sup>

<sup>1</sup> Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>2</sup> Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>3</sup> Division of Cardiovascular Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>4</sup> Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA

- 
- <sup>5</sup> Division of General Internal Medicine and Public Health, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
- <sup>6</sup> 23andMe, Inc., Mountain View, CA, USA
- <sup>7</sup> Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands
- <sup>8</sup> Department Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands
- <sup>9</sup> Department of Psychiatry, University of California San Diego, La Jolla, CA, USA
- <sup>10</sup> Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA
- <sup>11</sup> Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands
- <sup>12</sup> APH (Amsterdam Public Health) Institute, Amsterdam, The Netherlands
- <sup>13</sup> Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA
- <sup>14</sup> Departments of Medicine and Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA