Archival Report

Biological Psychiatry

Polygenic Liability to Depression Is Associated With Multiple Medical Conditions in the Electronic Health Record: Phenome-wide Association Study of 46,782 Individuals

Yu Fang, Lars G. Fritsche, Bhramar Mukherjee, Srijan Sen, and Leah S. Richmond-Rakerd

ABSTRACT

BACKGROUND: Major depressive disorder (MDD) is a leading cause of disease-associated disability, with much of the increased burden due to psychiatric and medical comorbidity. This comorbidity partly reflects common genetic influences across conditions. Integrating molecular-genetic tools with health records enables tests of association with the broad range of physiological and clinical phenotypes. However, standard phenome-wide association studies analyze associations with individual genetic variants. For polygenic traits such as MDD, aggregate measures of genetic risk may yield greater insight into associations across the clinical phenome.

METHODS: We tested for associations between a genome-wide polygenic risk score for MDD and medical and psychiatric traits in a phenome-wide association study of 46,782 unrelated, European-ancestry participants from the Michigan Genomics Initiative.

RESULTS: The MDD polygenic risk score was associated with 211 traits from 15 medical and psychiatric disease categories at the phenome-wide significance threshold. After excluding patients with depression, continued associations were observed with respiratory, digestive, neurological, and genitourinary conditions; neoplasms; and mental disorders. Associations with tobacco use disorder, respiratory conditions, and genitourinary conditions persisted after accounting for genetic overlap between depression and other psychiatric traits. Temporal analyses of time-at-first-diagnosis indicated that depression disproportionately preceded chronic pain and substance-related disorders, while asthma disproportionately preceded depression.

CONCLUSIONS: The present results can inform the biological links between depression and both mental and systemic diseases. Although MDD polygenic risk scores cannot currently forecast health outcomes with precision at the individual level, as molecular-genetic discoveries for depression increase, these tools may augment risk prediction for medical and psychiatric conditions.

https://doi.org/10.1016/j.biopsych.2022.06.004

Lifespan gains have stalled in many countries, since even before the COVID-19 pandemic. Between 2011 and 2017, multiple developed nations, including the United States, had lower gains in life expectancy than in prior years (1). This concerning trend amplifies the need to detect groups at high risk for poor health outcomes and implement early prevention strategies.

Depression is a strong target for prevention efforts to extend life expectancy. Depression is a leading cause of disease-associated disability (2), and the burden of disease due to depression encompasses not only psychiatric comorbidity but also medical comorbidity. Major depressive disorder (MDD) is associated with medical conditions including cardiovascular disease, diabetes, chronic lung disease, and cancer (3–6). These associations may arise through multiple mechanisms: depression may increase risk for poor physical health [e.g., via inflammation (7)], individuals with depression may engage in poor health behaviors and have difficulty accessing good medical care (8,9), and medical conditions may act as stressors to precipitate depression (10,11). Associations between depression and medical diseases also reflect shared genetic influences. Approximately 30% to 40% of the variation in major depression is attributable to genetic factors (12), some of which may also influence risk for medical conditions including cardiovascular disease (13–16), though findings are mixed (17,18).

Integration of molecular-genetic tools with electronic health records (EHRs) has enabled new approaches to test gene-disease associations. EHRs facilitate tests of association with the broad spectrum of physiological and clinical phenotypes. Standard phenome-wide association studies (PheWASs) analyze associations with individual genetic variants. For polygenic traits such as MDD, aggregate measures of genetic risk–including polygenic risk scores (PRSs)–may yield greater insight into associations across the clinical phenome (19–24). Such prediction efforts could yield at least 3 benefits.

ISSN: 0006-3223

First, testing associations between MDD-PRSs and the phenome can differentiate shared and distinct pathways of genetic association across diseases. Second, quantifying relations between MDD-PRSs and diagnoses from real-world clinical settings can help determine to what degree molecular-genetic measures might be used for risk prediction and stratification clinically (25). Third, depression and other mental disorders tend to emerge in adolescence and young adulthood, while noninfectious physical diseases and neurodegenerative conditions peak later in life (26). If genetic predisposition to depression is associated with physical diseases, early screening and prevention among individuals at high genetic risk for depression might benefit later-life—not just early-life—health. In this study, we leveraged genome-wide polygenic scores derived from a genome-wide meta-analysis of MDD (27) and EHR data on 46,782 individuals to test for associations between genetic liability to depression and the medical and psychiatric phenome.

To differentiate PRS associations driven by the primary trait diagnosis from independent associations arising through genetic risk, we conducted an exclusion PRS PheWAS (28) removing patients with a depression diagnosis and a second exclusion PheWAS removing patients with any psychological or developmental/behavioral disorder [to account for genetic overlap between depression and other psychiatric disorders (29,30)]. We leveraged the EHRs' time-stamped data to analyze the temporal order of diagnoses associated with the MDD-PRS in the first exclusion PheWAS, to determine whether they tended to predate or occur after depression. Different temporal sequences can provide insight into whether gene-disease associations might reflect processes beyond genetic overlap (e.g., causal pathways or diagnostic and social factors that influence the timing of diagnoses).

METHODS AND MATERIALS

Michigan Genomics Initiative Cohort

The Michigan Genomics Initiative (MGI) is a longitudinal biorepository effort enriched for patient genome-wide data and electronic health information (28). Participants were recruited through Michigan Medicine while awaiting diagnostic or interventional procedures either during a preoperative visit prior to the procedure or on the day of procedure that required anesthesia. Opt-in written informed consent was obtained. In addition to coded biosamples and protected secure health information, participants understood that all EHRs, claims, and national data-sources linkable to the participant may be incorporated into the MGI databank. Each participant donated a blood sample for genetic analysis and underwent baseline vital signs and a comprehensive history and physical assessment. Data were collected according to Declaration of Helsinki principles. Written consent forms and study protocols were reviewed and approved by the University of Michigan Medical School Institutional Review Board. Here, we report results obtained from 46,782 genotyped, unrelated European-ancestry samples with available integrated EHR data, recruited through the standard MGI study and 3 substudies contributing to the MGI databank: Metabolism, Endocrinology and Diabetes, Michigan Predictive Activity and Clinical Trajectories Study, and Mental Health Biobank.

Genotyping, Quality Control, and Imputation

DNA from 56,984 blood samples was genotyped on 2 batches of customized Illumina Infinium CoreExome-24 bead arrays (Illumina, Inc.) (n_1 = 19,931, n_2 = 37,053) [see (28,31,32) for array descriptions]. Complete details of quality-control procedures identifying and reducing potential batch effects are described in (28,33). Principal components (PCs) and ancestry were estimated by projecting all genotyped samples into the space of the principal components of the Human Genome Diversity Project reference panel using PLINK (34,35). Only patients of inferred recent European descent were included. Pairwise kinship was assessed using KING (36). FastIndep was used to reduce the data to a maximal subset that contained no pairs of individuals with third- or closer-degree relationship (37). Additional genotypes were obtained using the Haplotype Reference Consortium reference panel using the Michigan Imputation Server (38) and included over 24 million imputed variants with $R^2 > 0.3$ and minor allele frequency $> 0.1\%$. After further restricting the data to subjects with complete diagnosis and age information, our analytic sample comprised 46,782 individuals.

Phenome Generation

The MGI phenome was based on ICD-9 and ICD-10 codes. A total of 32,706 ICD codes were aggregated to PheWAS trait codes (phecodes) using the PheWAS R package (39). ICD-9 codes were aggregated using ICD-9 Phecode Map 1.2 (40–42). ICD-10 codes were aggregated using ICD-10-CM Phecode Map 1.2 beta (43,44). Phecodes are organized hierarchically, with parent phecodes comprising subgroups of child phecodes that enable ascertainment of traits at different levels of detail (45). For each phecode, we identified cases as individuals who had at least 1 assignment of that code in their records, and controls as the remaining individuals who did not have any assignment of overlapping phecodes defined by the exclusion criteria (41–44). For example, we excluded cases with overlapping phecodes ranging from 295 to 306.99 when generating controls for MDD (phecode 296.22). To minimize differences in age and sex distributions or extreme case-control ratios and to reduce computational burden, for each trait, we matched up to 10 controls to each case using the R package MatchIt (39,46). Nearest-neighbor matching was applied for age and PC1-4 (Mahalanobis-metric matching; matching-window caliper/width = 0.25 SDs) and exact matching was applied for sex and genotyping batch. We excluded phecodes with ≤50 cases. In total, 1814 case-control studies were generated, of which 1685 with >50 cases were used for analyses.

PRS Construction

We obtained genome-wide association study (GWAS) summary statistics for MDD from a meta-analysis of the Psychiatric Genomics Consortium MDD phase 2, UK Biobank, and 23andMe containing 246,363 cases and 561,190 controls of European ancestry (27). We reduced these summary statistics from 10.5 million single nucleotide polymorphisms (SNPs) to 1.1 million nonambiguous SNPs reported with minor allele frequency >1% and overlapping with Haplotype Reference

Consortium-imputed MGI data and the European-ancestry linkage disequilibrium reference panel of PRS-CS (https://github.com/getian107/PRScs).

For PRS calculation we used PRS-CS-auto (47), a method that uses a high-dimensional Bayesian regression framework with continuous shrinkage priors that infers posterior effect sizes of SNPs while avoiding overfitting and enables automatic learning of the global shrinkage parameter from GWAS summary statistics. Each participant's MDD-PRS was then calculated using the R package Rprs (https://github.com/statgen/Rprs), which takes variant lists and weights from PRS-CS output and the allele dosages of the Haplotype Reference Consortium-imputed MGI data.

PheWAS Analyses

For each of the 1685 phenome disease traits, we fit a Firth bias-corrected logistic regression model adjusting for age, sex, genotyping batch, and recruitment study:

$$\text{Logit}(P[\text{Disease} = 1 \mid \text{MDD-PRS, Age, Sex, Batch, Study, PC1-PC4}]) = \beta 0 + \beta_{PRS}\,\text{MDD-PRS} + \beta_{Age}\text{Age} + \beta_{Sex}\text{Sex} + \beta_{Batch}\text{Batch} + \beta_{Study}\text{Study} + \beta_{PC}\,(\text{PC1}, .., \text{PC4}),$$

where β_{PRS} estimates the association between the MDD-PRS and the trait. The p value corresponding to tests of H_0: $\beta_{PRS} = 0$ was used to estimate the significance of this association. Bonferroni-correction resulted in a significance threshold of $p < 3.2 \times 10^{-5}$ ($\simeq 0.05/1685$).

To help differentiate MDD-PRS associations driven by shared genetics from those driven by the primary depression diagnosis, we performed an exclusion PRS PheWAS (28) excluding patients with a depression diagnosis (phecode 296.2). To determine whether MDD-PRS associations were explained by shared genetic architecture between depression and other psychiatric conditions, we performed an additional exclusion PheWAS excluding patients with any psychological or developmental/behavioral disorder (phecodes 295–306.99).

Statistical analyses were performed using R version 4.1.1 (R Foundation for Statistical Computing; https://www.r-project.org/).

Temporal Analysis

We conducted a temporal analysis of diseases that had the strongest positive associations with the MDD-PRS in the depression-specific exclusion PheWAS, using patient-level data on date of first diagnosis within the full PheWAS sample. For each disease (among patients with diagnoses of both MDD and that disease), we plotted horizontal bars indicating each patient's time of first trait diagnosis relative to time of first MDD diagnosis, in months, and counted the total numbers of patients who received their first non-MDD diagnosis later than, within 1 month of, or earlier than their first MDD diagnosis.

RESULTS

Sample Characteristics

This study included 46,782 unrelated individuals of European ancestry (52.3% female, mean age = 56.7 years [SD = 16.4] (Table 1). For MDD (phecode 296.22), there were 13,850 cases and 20,853 matched controls (Table 1). Its parent trait, depression (phecode 296.2), had 14,833 cases and 20,856 matched controls. The MDD-PRS explained 1.2% to 2.2% of the variance in MDD (pseudo-R^2 = 0.012 [McFadden], 0.022 [Nagelkerke]).

MDD-PRS PheWAS

The MDD-PRS was associated with 211 traits from 15 disease categories at the Bonferroni-corrected significance threshold $p < 3.2 \times 10^{-5}$ (Figure 1; Table S1 in Supplement 2; Figure S1 in Supplement 1). Of the top 50 traits, 18 were in the mental disorder category and 32 were in nonpsychiatric categories. MDD showed the strongest association (odds ratio [OR] = 1.31 [1.28–1.35], $p = 1.7 \times 10^{-120}$). Of the nonpsychiatric diseases, the strongest associations in the digestive, neurological, respiratory, musculoskeletal, circulatory-system, and symptoms categories were observed for gastroesophageal reflux disease (OR = 1.15 [1.13–1.17], $p = 2.8 \times 10^{-41}$), other headache syndromes (OR = 1.15 [1.12–1.17], $p = 2.2 \times 10^{-28}$), chronic bronchitis (OR = 1.19 [1.15–1.23], $p = 1.3 \times 10^{-24}$), osteoarthrosis, not otherwise specified (OR = 1.10 [1.08–1.12], $p = 2.3 \times 10^{-19}$), nonspecific chest pain (OR = 1.09 [1.06–1.11], $p = 8.8 \times 10^{-15}$), and back pain (OR = 1.10 [1.07–1.12], $p = 1.4 \times 10^{-19}$), respectively. The MDD-PRS was negatively associated with benign neoplasm of lymph nodes (OR = 0.94 [0.91–0.97], $p = 1.2 \times 10^{-5}$) in the neoplasms category (Table S1 in Supplement 2; Figure S1B in Supplement 1).

Exclusion PRS PheWAS

We conducted an exclusion PheWAS excluding the 14,833 patients with a depression diagnosis (parent phecode 296.2, which includes the child phecode MDD [296.22]) (Table S1 in Supplement 2). An additional 117 phecodes were excluded because the number of cases dropped below 50, resulting in 1566 traits for analysis. The MDD-PRS was associated with 25 traits from 6 disease categories at the Bonferroni-corrected significance threshold $p < 3.2 \times 10^{-5}$ (Figure 2; Table S1 in Supplement 2; Figure S1 in Supplement 1). These categories included mental disorders; digestive, neurological, respiratory, and genitourinary diseases; and neoplasms. For the 25 traits, the effect sizes (unstandardized regression coefficients) for the exclusion-PheWAS associations were on average 73% of the effect sizes for the full-PheWAS associations (range = 0.52–1.18) (Table S1 in Supplement 2). Figure S1 in Supplement 1 shows the change in OR.

Table 1. Sample Characteristics (N = 46,782)

Characteristic	Analytic Sample
Females, *n* (%)	24,454 (52.3%)
Age, Years, Mean (SD)	56.7 (16.4)
Number of Unique ICD Codes	32,706
Number of Unique Phecodes	1685
Median Number of Days Between First and Last Visit	2261
Individuals With Major Depressive Disorder, *n* (%)	13,850 (29.6%)

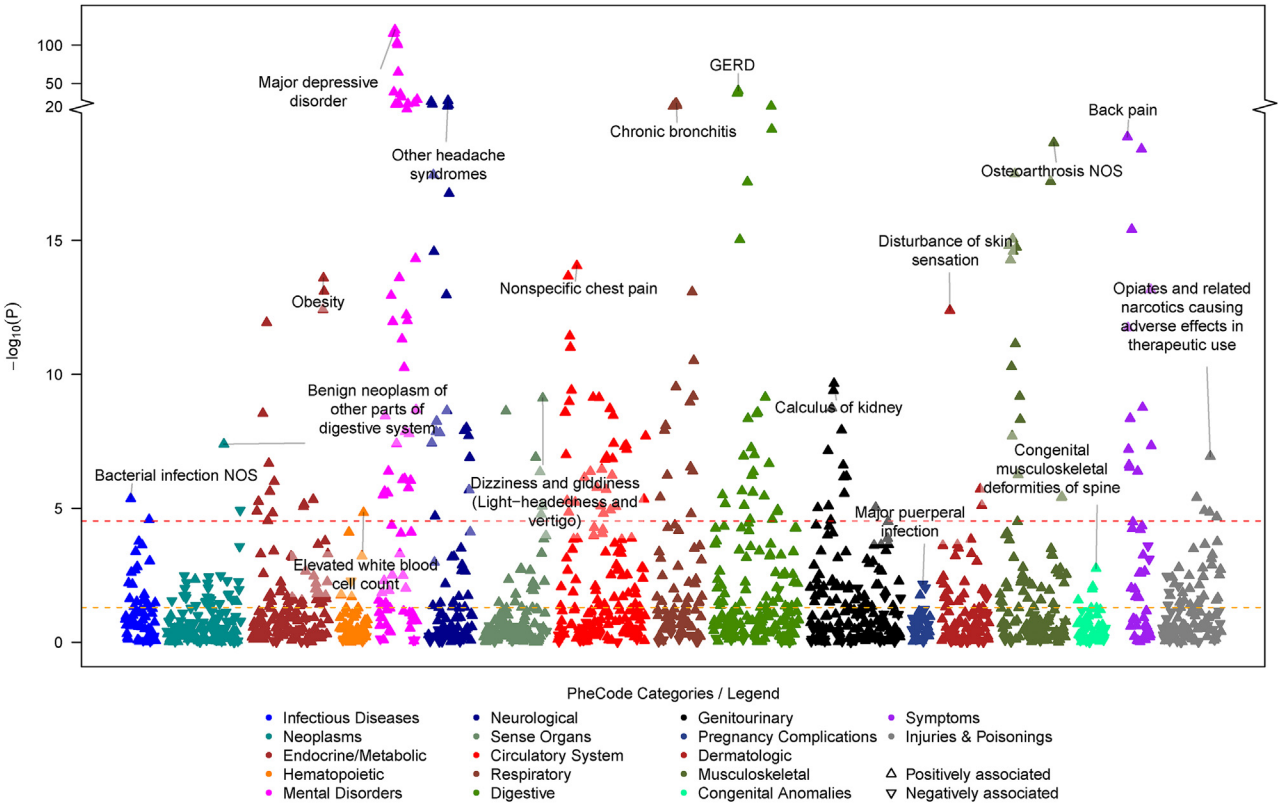

Figure 1. Phenome-wide association of the major depressive disorder polygenic risk score. The figure shows the $-\log_{10}$ (p values) for associations between the major depressive disorder polygenic risk score and the phecodes for 1685 disease traits, grouped by 17 color-coded categories. Directional triangles indicate the direction of association. The yellow dashed line indicates a significance threshold of $p = .05$. The red dashed line indicates the Bonferroni-corrected phenome-wide significance threshold ($-\log_{10}$ [p value] = 4.5). Traits with the strongest association within each category are labeled. GERD, gastroesophageal reflux disease; NOS, not otherwise specified.

A total of 8 mental disorder traits remained significantly associated with the MDD-PRS in the exclusion PheWAS: anxiety disorders (OR = 1.15 [1.12–1.19], $p = 1.9 \times 10^{-19}$), anxiety disorder (OR = 1.15 [1.12–1.19], $p = 3.1 \times 10^{-18}$), generalized anxiety disorder (OR = 1.16 [1.09–1.23], $p = 7.5 \times 10^{-7}$), tobacco use disorder (OR = 1.09 [1.06–1.13], $p = 6.1 \times 10^{-8}$), posttraumatic stress disorder (OR = 1.53 [1.29–1.82], $p = 8.2 \times 10^{-7}$), bipolar disorder (OR = 1.40 [1.21–1.61], $p = 2.9 \times 10^{-6}$), dysthymic disorder (OR = 1.33 [1.17–1.51], $p = 1.1 \times 10^{-5}$), and substance addiction and disorders (OR = 1.13 [1.07–1.19], $p = 2.1 \times 10^{-5}$) (Table S1 in Supplement 2; Figure S1E in Supplement 1).

Of the nonpsychiatric diseases, the traits showing the strongest associations in the exclusion PheWAS were gastroesophageal reflux disease (OR = 1.09 [1.06–1.12], $p = 6.2 \times 10^{-11}$) in the digestive diseases, other headache syndromes (OR = 1.09 [1.05–1.12], $p = 1.4 \times 10^{-6}$) in the neurological diseases, chronic airway obstruction (OR = 1.12 [1.08–1.17], $p = 1.1 \times 10^{-8}$) in the respiratory diseases, calculus of kidney (OR = 1.10 [1.06–1.14], $p = 5.9 \times 10^{-8}$) in the genitourinary diseases, and benign neoplasm of skin (OR = 0.94 [0.91–0.96], $p = 2.3 \times 10^{-6}$) in the neoplasms. Two other neoplasm traits—benign neoplasms of lymph nodes and unspecified sites—were also negatively associated with the MDD-PRS in the exclusion PheWAS (Table S1 in Supplement 2; Figure S1B in Supplement 1). The largest proportion of negative associations within each disease category (regardless of statistical significance) was observed for the neoplasms (81.9%) (Table S2 in Supplement 1).

We conducted a second exclusion PheWAS excluding the 26,329 patients with any psychological or developmental/behavioral disorder (43 phecodes: 295–315.99). After further excluding phecodes with <50 cases, 1355 traits were available for analysis. The MDD-PRS was associated with 7 traits from 3 disease categories at the Bonferroni-corrected significance threshold $p < 3.2 \times 10^{-5}$, all of which were also significant in the depression-specific exclusion PheWAS: tobacco use disorder (OR = 1.11 [1.06–1.16], $p = 2.2 \times 10^{-6}$) in the mental disorders; chronic airway obstruction (OR = 1.14 [1.08–1.20], $p = 8.4 \times 10^{-7}$), asthma (OR = 1.10 [1.06–1.15], $p = 5.6 \times 10^{-6}$), chronic bronchitis (OR = 1.15 [1.08–1.22], $p = 6.4 \times 10^{-6}$), and obstructive chronic bronchitis (OR = 1.15 [1.08–1.22], $p = 1.3 \times 10^{-5}$) in the respiratory diseases; and calculus of kidney (OR = 1.11 [1.06–1.16], $p = 4.6 \times 10^{-6}$) and urinary calculus (OR = 1.10 [1.05–1.15], $p = 1.7 \times 10^{-5}$) in the genitourinary diseases (effect sizes for significant traits = 114% of those in the depression-specific exclusion PheWAS, on average [range = 1.09–1.22]) (Figure S2 in Supplement 1). Gastroesophageal reflux disease continued to show the strongest association within the digestive diseases, though the

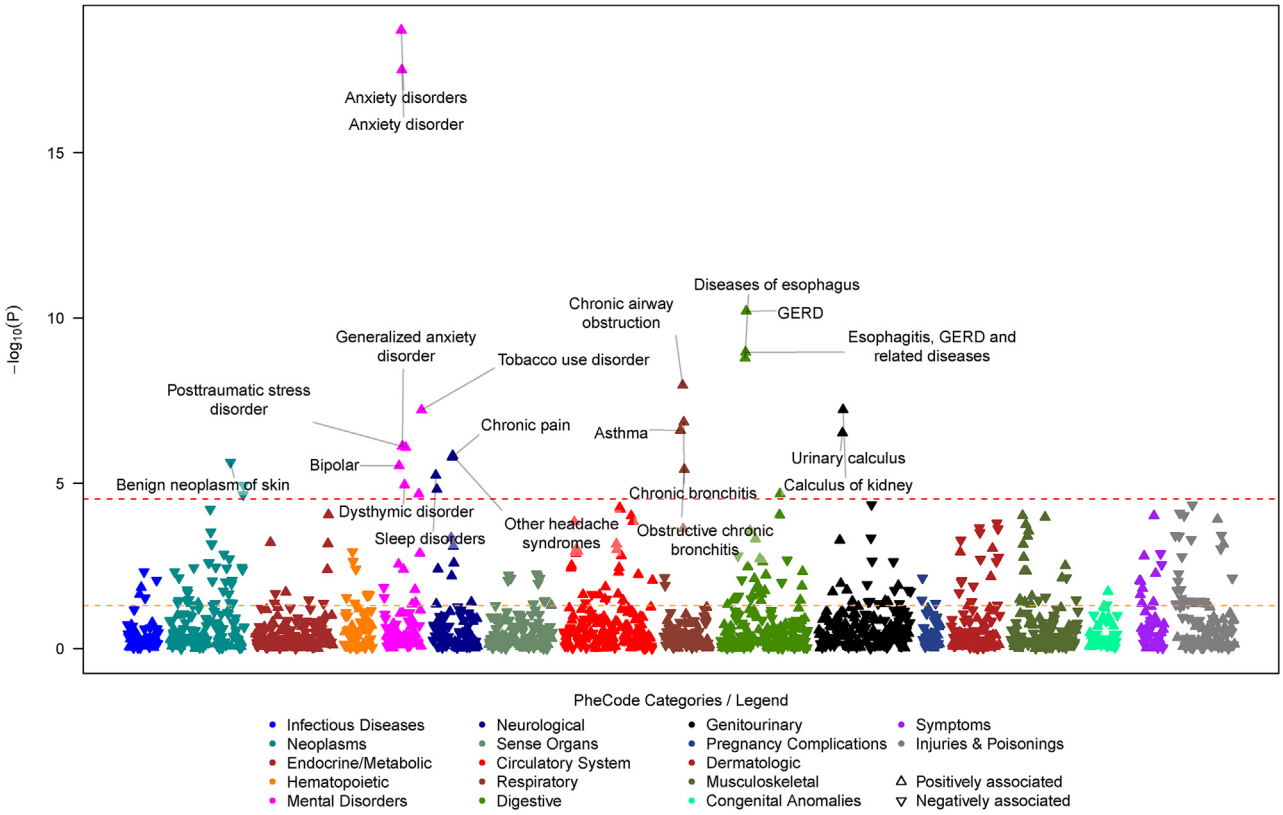

Figure 2. Phenome-wide association of the major depressive disorder polygenic risk score, excluding patients with a depression diagnosis. The figure shows the $-\log_{10}$ (*p* values) for associations with the major depressive disorder polygenic risk score, after removing patients with depression (phecode 296.2, which includes major depressive disorder [phecode 296.22]). Phecodes for which the number of cases dropped below 50 (n = 117) were also excluded, resulting in a total of 1566 disease traits for analysis. The traits are grouped by 17 color-coded categories. Directional triangles indicate the direction of association. The yellow dashed line indicates a significance threshold of p = .05. The red dashed line indicates the Bonferroni-corrected significance threshold ($-\log_{10}$ [*p* value] = 4.5). The 20 traits with the strongest associations are labeled. GERD, gastroesophageal reflux disease.

association did not reach phenome-wide significance (OR = 1.07 [1.04–1.11], $p = 4.1 \times 10^{-5}$, effect size = 83% of that in the depression-specific exclusion PheWAS) (Figure S2 in Supplement 1).

Temporal Analysis

We analyzed the temporal associations between MDD and traits that showed evidence of genetic overlap with depression. Within the full PheWAS sample, we selected the 12 top parent phecodes for traits with the strongest positive associations with the MDD-PRS in the depression-specific exclusion PheWAS and plotted their time at first diagnosis relative to the time at first diagnosis for MDD (Figure 3). Several diagnoses tended to disproportionately succeed or precede depression; nearly 3 times as many patients received their first chronic pain diagnosis and over twice as many received their first substance addiction and disorders diagnosis after their first MDD diagnosis than before (Figure 3C, F). By contrast, over twice as many patients received their first asthma diagnosis before their first MDD diagnosis (Figure 3H). The distributions of time at first diagnosis for the traits relative to MDD deviated significantly from the average distributions of time at first diagnosis for all other nonoverlapping phecodes relative to MDD (Table S3 in Supplement 2).

The patterns we observed might reflect differences in age of onset. However, age at first diagnosis for the 12 traits was not associated with the proportion of cases diagnosed before or after MDD (p = .58 and .95, respectively). In addition, for 11 of the 12 traits, the difference in median age at first diagnosis between the trait and MDD was shorter among patients with both diagnoses, relative to those with only MDD or the trait. For all 12 traits, patients who also had an MDD diagnosis showed an earlier age at first diagnosis of the trait than those without an MDD diagnosis (Table S3 in Supplement 2). These findings suggest that factors beyond age of onset may help to explain these diseases' temporal sequencing.

We evaluated whether 2 psychiatric conditions that share genetic architecture with depression (anxiety disorder and attention-deficit/hyperactivity disorder [ADHD]) showed a similar pattern to MDD in their temporal sequencing with traits with evidence of genetic overlap with depression (those in Figure 3). Anxiety disorder and ADHD tended to show a similar pattern to MDD; across all traits (excluding anxiety disorder), 79% (±2%) of patients, on average, received both their first anxiety disorder and their first MDD diagnosis before, after, or

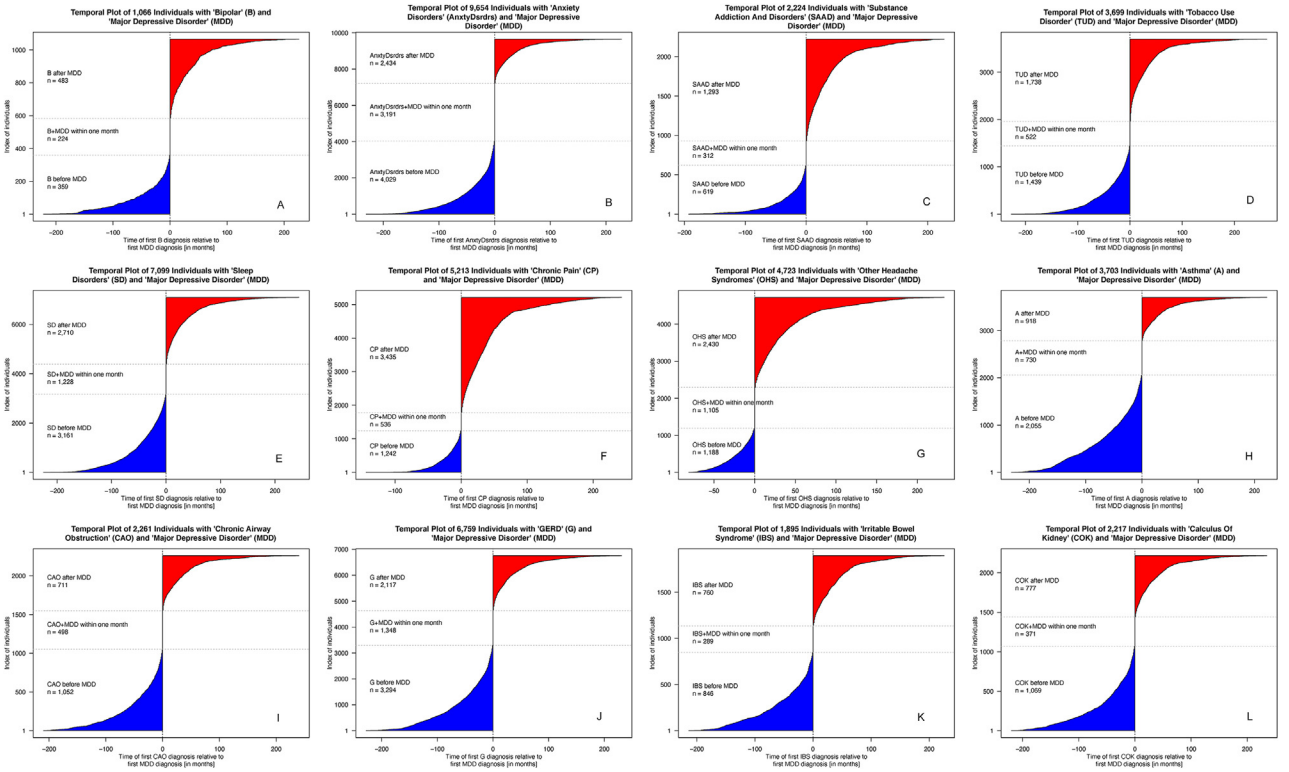

Figure 3. Temporal order of diagnoses. The plot shows the number of patients who received their first non–major depressive disorder diagnosis earlier (in blue) and later (in red) than their first major depressive disorder diagnosis, distributed by relative time in months. A modest number of individuals who were included in the phenome-wide association study analyses are excluded from this plot because they were missing information on date of diagnosis. **(A)** Bipolar disorder. **(B)** Anxiety disorders. **(C)** Substance addiction and disorders. **(D)** Tobacco use disorder. **(E)** Sleep disorders. **(F)** Chronic pain. **(G)** Other headache syndromes. **(H)** Asthma. **(I)** Chronic airway obstruction. **(J)** Gastroesophageal reflux disease. **(K)** Irritable bowel syndrome. **(L)** Calculus of kidney.

within 1 month of the other trait. For ADHD, this number was 72% (±3%).

DISCUSSION

In this PheWAS of over 46,000 individuals, we found that a genome-wide polygenic score for MDD was associated with multiple medical and psychiatric conditions in the EHR. After excluding patients with depression, continued associations were observed with respiratory, digestive, neurological, and genitourinary conditions; neoplasms; and mental disorders. Associations with tobacco use disorder, respiratory conditions, and genitourinary conditions persisted after accounting for genetic overlap between depression and other psychiatric traits. Temporal analyses indicated variation across conditions in their distributions of time at first diagnosis relative to MDD.

Our analysis has strengths. First, we used a well-powered study cohort and generated the MDD-PRS using a method with a flexible modeling assumption that adaptively approximates the true effect-size distribution (48). Second, our exclusion-PheWAS approach allowed us to isolate genetically mediated associations from those driven by depression or related psychiatric diagnoses or bias in the EHR data. Although this approach reduces sample size, it is statistically preferable to controlling for the primary trait (28). Third, EHRs' temporal data have received limited attention in prior psychiatric PRS PheWASs. Leveraging these data enabled us to further interrogate gene–disease associations.

Associations were observed across both exclusion PheWASs between the MDD-PRS and respiratory diseases including asthma, chronic airway obstruction (also termed chronic obstructive pulmonary disease), chronic bronchitis, and obstructive chronic bronchitis. Prior work has demonstrated prospective relations between mental disorders including depression with chronic obstructive pulmonary disease in representative cohorts (49) and complete populations (50), and previous MDD-PRS PheWASs have also identified associations with respiratory conditions (22,23), with evidence that MDD may be a causal risk factor for asthma (22). Our exclusion analyses suggest that associations between MDD and several respiratory diseases are at least in part genetically driven and not driven solely by the primary depression diagnosis or genetically related psychiatric conditions. Our exclusion PheWAS also revealed associations with diseases that are in nonrespiratory categories but are related to the respiratory system, including tobacco use disorder, which has been identified as a top association in a previous MDD-PRS PheWAS (23).

Consistent with prior PRS PheWASs of MDD (22) and schizophrenia (23,25), our full PheWAS uncovered significant associations with genitourinary conditions. Many of these associations became nonsignificant in the exclusion PheWAS,

indicating that they may arise as a function of the depression diagnosis. For instance, genitourinary problems are a side effect of some antidepressants (51,52). Exceptions were calculus of kidney and urinary calculus, suggesting genetic correlations with MDD. Though requiring replication, these novel associations suggest variation in the mechanisms linking genetic liability to depression with different urinary-system disorders.

Prior quantitative-genetic and molecular-genetic studies have identified anxiety and bipolar disorders as among the psychiatric disorders with the strongest genetic correlations with MDD (53–55). Anxiety disorders and bipolar disorder emerged as top hits in our depression-specific exclusion PheWAS, corroborating prior findings and showing that these genetic correlations extend to diagnoses in the clinical phenome.

Negative associations emerged with benign neoplasms of lymph nodes, skin, and unspecified sites in our depression-specific exclusion PheWAS. Furthermore, when considering all associations in the depression-specific exclusion analysis—regardless of statistical significance—the largest proportion of negative associations was in the neoplasms category. However, neoplasm associations did not remain significant after accounting for genetic overlap between depression and other psychiatric traits. A prior PheWAS (23) found a negative association between an ADHD-PRS and benign neoplasm of skin, providing additional evidence that the neoplasm associations we observed may be attributable at least in part to genetic influences shared between MDD and related psychiatric conditions, including ADHD. Systematic reviews and meta-analyses of depression and cancer indicate a positive association in the population (56,57). Our results suggest that depression and neoplasms may be positively associated at the phenotypic level, but—similar to the negative genetic association between neurodegenerative disease and cancer (58)—the genetic predisposition to depression or related psychiatric conditions may predispose to less cancer. More work is needed to clarify these associations' robustness and their specificity across psychiatric conditions.

Relative to prior MDD-PRS EHR PheWASs (21–24), we observed associations across a broader range of disease categories in our full PheWAS, including novel associations in 2 categories (neoplasms and neurological disorders) that remained significant in our depression-specific exclusion PheWAS. This may reflect differences in sample size or variants included in the PRS; prior MDD-PRS EHR PheWASs have used smaller sample sizes (21,23,24) or selected only the top GWAS variants to construct their PRS (22), which may introduce information loss (59).

The temporal ordering of diagnoses provides insights into potential prevention opportunities. We found that a disproportionate number of individuals were first diagnosed with depression prior to chronic pain and substance addiction and disorders, suggesting that some of the genetic risk for pain and addiction may operate through depressive mechanisms, and ameliorating depression might also ameliorate risk for these disorders. Conversely, a disproportionate number of individuals were first diagnosed with asthma prior to depression, suggesting that mitigating asthma symptoms may also mitigate risk for depression. Our data cannot resolve causality, and other mechanisms may help to explain disorders' temporal sequencing, including variations in diagnostic criteria, greater hesitancy to seek treatment for psychiatric conditions than physical health conditions because of concerns about stigma, and age of onset differences (although analyses suggested these differences could not fully explain observed temporal patterns). Depression and temporally antecedent or subsequent conditions may also be indicators of a broader disease progression process. Our findings suggest causal hypotheses that could be tested in future Mendelian-randomization studies and randomized clinical trials.

These findings have additional implications. First, results of our exclusion PheWAS indicate that polygenic liability to MDD predicts risk for medical conditions regardless of whether individuals meet depression diagnostic thresholds. This suggests that molecular-genetic measures of depression risk may be informative for all patients. Furthermore, depression and other mental disorders tend to onset in adolescence and young adulthood, while noninfectious physical diseases peak in mid to late life (26). Early screening among young people at high genetic risk for depression might yield insights into their longer-term risk for poor physical health. However, it is important to note that the MDD-PRS is not currently sensitive or specific enough to forecast health outcomes with precision at the individual level. MDD-PRSs will need to differentiate cases from controls with greater accuracy before they can be used clinically.

Second, our findings can inform etiologic research. GWASs of depression and other phenotypes prioritize selection of cases who represent the disorder of interest, and often apply exclusion criteria for other conditions. However, our results suggest that MDD genetic-discovery efforts may be enhanced by increasing representation within GWAS samples of individuals also diagnosed with associated health conditions, in particular, conditions for which associations are genetically driven. The predictive power of PRSs for depression [and for other psychiatric disorders (23)] may be increased by including, in GWAS samples, individuals who represent a broader range of associated diseases.

Third, our results highlight opportunities for interdisciplinary research. We have shown that molecular-genetic risk for depression is associated with physical-disease diagnoses in clinical records. An important next step is to identify the phenotypes that connect polygenic liability to depression to poor physical health across development. Such work will require integrating molecular-genetic tools within prospective cohort studies (60) and collaborative cross talk between researchers and clinicians in developmental science, genomics, psychiatry, and geriatric medicine.

Fourth, of the 1685 phenotypes analyzed, only 12.5% were significant in the full PheWAS, substantially lower than the number of phenotypes with which depression co-occurs. This reinforces the importance of identifying environmental as well as genetic mechanisms of comorbidity. It should be noted, however, that the number of cases varied substantially across conditions (full PheWAS: range = 51–23,317). Thus, our study cannot rule out genetic associations for rare conditions for which analyses may have been underpowered.

We acknowledge limitations. First, analyses were restricted to European-ancestry participants because of limited GWAS summary statistics for other populations. GWASs conducted in

more diverse samples will facilitate polygenic prediction among non-European-ancestry individuals. Second, the MGI is not a population-based sample. However, its medical-center-based ascertainment strategy is a compromise for access to EHR data. Although MGI participants are on average less healthy than the general population, this introduces a statistical-power advantage for PheWASs because the cohort is enriched for medical-disease cases (32). Third, potential biases in medical record data include phenotype misclassification and availability of data for only part of the life span. The EHRs in this study comprised a median of 6.2 years of data per person; we could not ascertain diagnoses made outside this window or in different health systems. Fourth, although the MGI EHRs integrate information about depression diagnoses made across primary care, outpatient treatment, and inpatient treatment, our analysis does not include depression diagnoses among individuals who are not assessed for depression during a health care contact. Thus, our exclusion analyses likely did not rule out all individuals with depression.

Polygenic liability to MDD provides a window into both mental and physical health. Our results can inform MDD genetic discovery efforts as well as developmental and epidemiologic research linking genomic risk for depression to physical diseases. MDD-PRSs cannot yet accurately forecast health outcomes at the individual level. However, as molecular-genetic discoveries for depression increase, these tools may become a useful component of risk prediction for both medical and psychiatric conditions.

ACKNOWLEDGMENTS AND DISCLOSURES

This study was supported by the National Institute of Mental Health (Grant No. MH101459 [to SS]) and the National Institute of Child Health and Human Development (Grant No. HD065563 through the Duke Population Research Center [to LSR-R]).

We thank the Michigan Genomics Initiative participants; Precision Health at the University of Michigan, the University of Michigan Medical School Central Biorepository, and the University of Michigan Advanced Genomics Core for providing data and specimen storage, management, processing, and distribution services; and the Center for Statistical Genetics in the Department of Biostatistics at the School of Public Health for genotype data curation, imputation, and management in support of the research reported in this publication.

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Michigan Neuroscience Institute (YF, SS), Department of Biostatistics (LGF, BM), Rogel Cancer Center (LGF, BM), Center for Statistical Genetics (LGF, BM), Department of Epidemiology (BM), Department of Psychiatry (SS), and Department of Psychology (LSR-R), University of Michigan, Ann Arbor, Michigan.

Address correspondence to Yu Fang, M.S.E., at yfang@umich.edu.

Received Oct 17, 2021; revised Apr 1, 2022; accepted Jun 2, 2022.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsych.2022.06.004.

REFERENCES

1. Crimmins EM (2021): Recent trends and increasing differences in life expectancy present opportunities for multidisciplinary research on aging. Nat Aging 1:12–13.
2. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators (2018): Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017 [published correction appears in Lancet 2019; 393:e44]. Lancet 392:1789–1858.
3. Patten SB, Williams JVA, Lavorato DH, Modgill G, Jetté N, Eliasziw M (2008): Major depression as a risk factor for chronic disease incidence: Longitudinal analyses in a general population cohort. Gen Hosp Psychiatry 30:407–413.
4. Rotella F, Mannucci E (2013): Depression as a risk factor for diabetes: A meta-analysis of longitudinal studies. J Clin Psychiatry 74:31–37.
5. Scott KM, Lim C, Al-Hamzawi A, Alonso J, Bruffaerts R, Caldas-de-Almeida JM, *et al.* (2016): Association of mental disorders with subsequent chronic physical conditions: World Mental Health Surveys from 17 countries. JAMA Psychiatry 73:150–158.
6. Whooley MA, de Jonge P, Vittinghoff E, Otte C, Moos R, Carney RM, *et al.* (2008): Depressive symptoms, health behaviors, and risk of cardiovascular events in patients with coronary heart disease. JAMA 300:2379–2388.
7. Kiecolt-Glaser JK, Glaser R (2002): Depression and immune function: Central pathways to morbidity and mortality. J Psychosom Res 53:873–876.
8. Penninx BWJH (2017): Depression and cardiovascular disease: Epidemiological evidence on their linking mechanisms. Neurosci Biobehav Rev 74:277–286.
9. Archer G, Kuh D, Hotopf M, Stafford M, Richards M (2020): Association between lifetime affective symptoms and premature mortality. JAMA Psychiatry 77:806–813.
10. Han KM, Kim MS, Kim A, Paik JW, Lee J, Ham BJ (2019): Chronic medical conditions and metabolic syndrome as risk factors for incidence of major depressive disorder: A longitudinal study based on 4.7 million adults in South Korea. J Affect Disord 257:486–494.
11. Tang PL, Wang HH, Chou FH (2015): A systematic review and meta-analysis of demoralization and depression in patients with cancer. Psychosomatics 56:634–643.
12. Sullivan PF, Neale MC, Kendler KS (2000): Genetic epidemiology of major depression: Review and meta-analysis. Am J Psychiatry 157:1552–1562.
13. Amare AT, Schubert KO, Klingler-Hoffmann M, Cohen-Woods S, Baune BT (2017): The genetic overlap between mood disorders and cardiometabolic diseases: A systematic review of genome wide and candidate gene studies. Transl Psychiatry 7:e1007.
14. Kendler KS, Gardner CO, Fiske A, Gatz M (2009): Major depression and coronary artery disease in the Swedish Twin Registry: Phenotypic, genetic, and environmental sources of comorbidity. Arch Gen Psychiatry 66:857–863.
15. Scherrer JF, Xian H, Bucholz KK, Eisen SA, Lyons MJ, Goldberg J, *et al.* (2003): A twin study of depression symptoms, hypertension, and heart disease in middle-aged men. Psychosom Med 65:548–557.
16. van Hecke O, Hocking LJ, Torrance N, Campbell A, Padmanabhan S, Porteous DJ, *et al.* (2017): Chronic pain, depression and cardiovascular disease linked through a shared genetic predisposition: Analysis of a family-based cohort and twin study. PLoS One 12:e0170653.
17. Khandaker GM, Zuber V, Rees JMB, Carvalho L, Mason AM, Foley CN, *et al.* (2020): Shared mechanisms between coronary heart disease and depression: Findings from a large UK general population-based cohort [published correction appears in Mol Psychiatry 2021; 26:3659–3661]. Mol Psychiatry 25:1477–1486.
18. Wium-Andersen MK, Villumsen MD, Wium-Andersen IK, Jørgensen MB, Hjelmborg Jvon B, Christensen K, Osler M (2020): The familial and genetic contribution to the association between depression and cardiovascular disease: A twin cohort study. Mol Psychiatry 26:4245–4253.
19. Shen X, Howard DM, Adams MJ, Hill WD, Clarke T-K, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, et al. (2020): A phenome-wide association and Mendelian randomisation study of polygenic risk for depression in UK Biobank. Nat Commun 11:2301.

20. Krapohl E, Euesden J, Zabaneh D, Pingault JB, Rimfeld K, von Stumm S, *et al.* (2016): Phenome-wide analysis of genome-wide polygenic scores. Mol Psychiatry 21:1188–1193.
21. McCoy TH, Castro VM, Snapper L, Hart K, Januzzi JL, Huffman JC, Perlis RH (2017): Polygenic loading for major depression is associated with specific medical comorbidity. Transl Psychiatry 7:e1238.
22. Mulugeta A, Zhou A, King C, Hyppönen E (2020): Association between major depressive disorder and multiple disease outcomes: A phenome-wide Mendelian randomisation study in the UK Biobank. Mol Psychiatry 25:1469–1476.
23. Kember RL, Merikangas AK, Verma SS, Verma A, Judy R, Regeneron Genetics Center, et al. (2021): Polygenic risk of psychiatric disorders exhibits cross-trait associations in electronic health record data from European ancestry individuals. Biol Psychiatry 89:236–245.
24. Dennis J, Sealock J, Levinson RT, Farber-Eger E, Franco J, Fong S, *et al.* (2021): Genetic risk for major depressive disorder and loneliness in sex-specific associations with coronary artery disease. Mol Psychiatry 26:4254–4264.
25. Zheutlin AB, Dennis J, Karlsson Linnér R, Moscati A, Restrepo N, Straub P, *et al.* (2019): Penetrance and pleiotropy of polygenic risk scores for schizophrenia in 106,160 patients across four health care systems. Am J Psychiatry 176:846–855.
26. Moffitt TE, Caspi A (2019): Psychiatry's opportunity to prevent the rising burden of age-related disease. JAMA Psychiatry 76:461–462.
27. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, *et al.* (2019): Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci 22:343–352.
28. Fritsche LG, Gruber SB, Wu Z, Schmidt EM, Zawistowski M, Moser SE, *et al.* (2018): Association of polygenic risk scores for multiple cancers in a phenome-wide study: Results from the Michigan Genomics Initiative. Am J Hum Genet 102:1048–1061.
29. Grotzinger AD, Mallard TT, Akingbuwa WA, Ip HF, Adams MJ, Lewis CM, *et al.* (2022): Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic, and molecular genetic levels of analysis. Nat Genet 54:548–559.
30. Cross-Disorder Group of the Psychiatric Genomics Consortium (2019): Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell 179:1469–1482.e11.
31. Surakka I, Fritsche LG, Zhou W, Backman J, Kosmicki JA, Lu H, *et al.* (2020): MEPE loss-of-function variant associates with decreased bone mineral density and increased fracture risk. Nat Commun 11:4093.
32. Zawistowski M, Fritsche LG, Pandit A, Vanderwerff B, Patil S, Schmidt EM, *et al.* (2021): The Michigan Genomics Initiative: A biobank linking genotypes and electronic clinical records in Michigan Medicine patients. medRxiv. https://doi.org/10.1101/2021.12.15.21267864.
33. Fritsche LG, Beesley LJ, VandeHaar P, Peng RB, Salvatore M, Zawistowski M, *et al.* (2019): Exploring various polygenic risk scores for skin cancer in the phenomes of the Michigan Genomics Initiative and the UK Biobank with a visual catalog: PRSWeb. PLoS Genet 15: e1008202.
34. Wang C, Zhan X, Bragg-Gresham J, Kang HM, Stambolian D, Chew EY, *et al.* (2014): Ancestry estimation and control of population stratification for sequence-based association studies. Nat Genet 46:409–415.
35. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, *et al.* (2008): Worldwide human relationships inferred from genome-wide patterns of variation. Science 319:1100–1104.
36. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM (2010): Robust relationship inference in genome-wide association studies. Bioinformatics 26:2867–2873.
37. Abraham KJ, Diaz C (2014): Identifying large sets of unrelated individuals and unrelated markers. Source Code Biol Med 9:6.
38. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, *et al.* (2016): A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48:1279–1283.
39. Carroll RJ, Bastarache L, Denny JC (2014): R PheWAS: Data analysis and plotting tools for phenome-wide association studies in the R environment. Bioinform 30:2375–2376.
40. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, *et al.* (2013): Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol 31:1102–1110.
41. PheWAS Resources: Phecode Map 1.2 with ICD-9 Codes. Available at: https://phewascatalog.org/phecodes. Accessed March 3, 2022.
42. Wei WQ, Bastarache LA, Carroll RJ, Marlo JE, Osterman TJ, Gamazon ER, *et al.* (2017): Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. PLoS One 12: e0175508.
43. Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, *et al.* (2019): Mapping ICD-10 and ICD-10-CM codes to phecodes: Workflow development and initial evaluation. JMIR Med Inform 7:e14325.
44. PheWAS Resources: Phecode Map 1.2 with ICD-10cm Codes (beta). Available at: https://phewascatalog.org/phecodes_icd10cm. Accessed March 3, 2022.
45. Bastarache L (2021): Using phecodes for research with the electronic health record: From PheWAS to PheRS. Annu Rev Biomed Data Sci 4:1–19.
46. Ho D, Imai K, King G, Stuart EA (2011): MatchIt: Nonparametric preprocessing for parametric causal inference. J Stat Softw 42:1–28.
47. Ge T, Chen CY, Ni Y, Feng Y-CA, Smoller JW (2019): Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun 10:1776.
48. Ma Y, Zhou X (2021): Genetic prediction of complex traits with polygenic scores: A statistical review. Trends Genet 37:995–1011.
49. Atlantis E, Fahey P, Cochrane B, Smith S (2013): Bidirectional associations between clinically relevant depression or anxiety and COPD: A systematic review and meta-analysis. Chest 144:766–777.
50. Richmond-Rakerd LS, D'Souza S, Milne BJ, Caspi A, Moffitt TE (2021): Longitudinal associations of mental disorders with physical diseases and mortality among 2.3 million New Zealand citizens. JAMA Netw Open 4:e2033448.
51. Carvalho AF, Sharma MS, Brunoni AR, Vieta E, Fava GA (2016): The safety, tolerability and risks associated with the use of newer generation antidepressant drugs: A critical review of the literature. Psychother Psychosom 85:270–288.
52. Trinchieri M, Perletti G, Magri V, Stamatiou K, Montanari E, Trinchieri A (2021): Urinary side effects of psychotropic drugs: A systematic review and metanalysis. Neurourol Urodyn 40:1333–1348.
53. Kendler KS, Aggen SH, Knudsen GP, Røysamb E, Neale MC, Reichborn-Kjennerud T (2011): The structure of genetic and environmental risk factors for syndromal and subsyndromal common DSM-IV axis I and all axis II disorders. Am J Psychiatry 168:29–39.
54. Smoller JW, Andreassen OA, Edenberg HJ, Faraone SV, Glatt SJ, Kendler KS (2019): Psychiatric genetics and the structure of psychopathology [published correction appears in Mol Psychiatry 2019; 24: 471. Mol Psychiatry 24:409–420.
55. Brain Consortium; Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, *et al.* (2018): Analysis of shared heritability in common disorders of the brain. Science 360(6395):eaap8757.
56. Jia Y, Li F, Liu YF, Zhao JP, Leng MM, Chen L (2017): Depression and cancer risk: A systematic review and meta-analysis. Public Health 149:138–148.
57. Wang YH, Li JQ, Shi JF, Que JY, Liu JJ, Lappin JM, *et al.* (2020): Depression and anxiety in relation to cancer incidence and mortality: A systematic review and meta-analysis of cohort studies. Mol Psychiatry 25:1487–1499.
58. Lanni C, Masi M, Racchi M, Govoni S (2021): Cancer and Alzheimer's disease inverse relationship: An age-associated diverging derailment of shared pathways. Mol Psychiatry 26:280–295.
59. Ware EB, Schmitz LL, Faul J, Gard A, Mitchell C, Smith JA, *et al.* (2017): Heterogeneity in polygenic scores for common human traits. bioRxiv. https://doi.org/10.1101/106062.
60. Belsky DW, Harden KP (2019): Phenotypic annotation: Using polygenic scores to translate discoveries from genome-wide association studies from the top down. Curr Dir Psychol Sci 28:82–90.