Increased depression genetic liability associates with inflammatory markers in multiple electronic health record systems

Lea Davis (lea.k.davis@gmail.com)  
Vanderbilt University Medical Center

Julia Sealock  
Vanderbilt University Medical Center  https://orcid.org/0000-0002-9346-3498

Younga Lee  
Massachusetts General Hospital  https://orcid.org/0000-0001-7517-6594

Arden Moscati  
Icahn School of Medicine at Mount Sinai

Sanan Venkatesh  
Icahn School of Medicine at Mount Sinai

Georgios Voloudakis  
Icahn School of Medicine at Mount Sinai  https://orcid.org/0000-0002-5729-632X

Peter Straub  
Vanderbilt University

Kritika Singh  
Vanderbilt University

Yen-Chen Feng  
Massachusetts General Hospital

Tian Ge  
Massachusetts General Hospital

Panos Roussos  
Icahn School of Medicine at Mount Sinai

Jordan W. Smoller  
Massachusetts General Hospital

Guanhua Chen  
University of Wisconsin-Madison

Article

Keywords: depression, inflammatory markers, genetic liability
Abstract

Although depression is a common disorder, its underlying biological basis remains poorly understood. The results of clinical lab tests available for research in electronic health records can be used to identify biomarkers that may reveal biological processes involved in the development of depression or may be markers of physiological changes due to depression. Here, we leveraged clinical laboratory tests and integrated biobank data to evaluate the relationship between genetic risk for depression and 315 routinely collected quantitative lab measures. Analyses across four health care systems (N = 382,452) robustly implicate increased white blood cell count as both a risk factor and consequence of depression diagnosis and revealed neutrophils, lymphocytes, and monocytes as the primary cell types responsible for this association. Our results highlight the importance of the immune system in the etiology of depression and motivate future development of clinical biomarkers and targeted treatment options for depression and its systemic effects.

Introduction

Depression is a common psychiatric disorder estimated to affect 264 million individuals worldwide \(^1\). Diagnostic criteria for depression include clinical evaluation of self-reported psychiatric symptoms, such as depressed mood, irritability, anhedonia, or suicidal thoughts. In addition to psychiatric effects, depression is associated with increased risk for cardiovascular disease \(^2\)–\(^4\), autoimmune disease \(^5\), and diabetes \(^6\)–\(^9\). The increased risk of peripheral diseases suggests the biology of depression is not limited to the brain; however, the causes and biologic effects of depression in the brain and the periphery remain poorly understood.

In a health care setting, laboratory ("lab") tests, are used to aid clinicians in diagnostic and treatment decision making, and tests that can accurately and reproducibly indicate a medical state are generally referred to as biomarkers \(^10\). To date there are no biomarkers for depression, however, consistent with the high number of common comorbidities, clinical depression is associated with a wide range of clinical labs, including increased pro-inflammatory cytokines \(^11\)–\(^14\), altered lipids \(^15\)–\(^17\), altered growth factors \(^18\)–\(^20\), and decreased brain-derived neurotrophic factor \(^21\)–\(^24\). For many of these physiological quantitative labs, the underlying biologic mechanisms are well understood. Thus further understanding the biological link between clinical depression and these labs can help identify the biological processes contributing to depression and could lead to the development of more informative biomarker panels to be used in risk assessment and treatment response.

Previous studies report a bidirectional relationship between depression and autoimmune disease \(^25\). Additionally, several immune biomarkers such as monocytes \(^26\)–\(^31\), neutrophil-lymphocyte ratio \(^14\), \(^32\), \(^33\), and C-reactive protein \(^14\), \(^31\), \(^34\) are increased in patients with depression compared to controls. However, most immune biomarker studies of depression have been limited in sample size and scope. Additionally, these
studies are often unable to control for potential confounders or determine the direction of association between depression and biomarkers.

Electronic health records (EHRs) store longitudinal information about the health and clinical care of individual patients, including lab results, medication records, diagnoses, and clinician notes. Biobanks that link EHRs to DNA provide an opportunity to analyze this clinical information along with genetic risk factors. Genetic risk for depression can be estimated using polygenic scores (PGS) which aggregate the small effects of thousands of loci across the genome into one score for each individual\textsuperscript{35}. Clinical lab data from EHRs has been historically difficult to use for research due to methodological issues, such as data entry errors, implausible recorded values, and changes in reference ranges and units over time\textsuperscript{36–38}. We recently developed and validated a scalable pipeline for quality control and harmonization of EHR-derived lab data (QualityLab), and analysis of cleaned labs that can enable large-scale genetic evaluation of laboratory values as biomarkers of disease (LabWAS)\textsuperscript{39}. In this work, we demonstrate that combining depression PGS with lab results stored in EHRs can robustly identify physiological processes that are affected by increased genetic liability to depression.

While independent biobanks can be used to discover associations, combining multiple health record systems through consortia can validate those discoveries in broader populations. The PsycheMERGE Network consists of investigators from institutions across the United States with the common goal of using EHRs and biobanks to advance the identification, biology, and treatment of psychiatric disorders\textsuperscript{40}. Here, we investigate the effect of polygenic risk for depression on clinically measured lab values leveraging data from four healthcare systems participating in the PsycheMERGE Network: Vanderbilt University Medical Center (VUMC), Massachusetts General Brigham (MGB), Mount Sinai Icahn School of Medicine (MSSM), and the Million Veteran Program (MVP).

**Results**

**LabWAS of Depression PGS**

Depression PGS were screened for associations with 315 clinical lab measurements using a lab-wide association scan (LabWAS)\textsuperscript{39} in VUMC’s biobank (N = 72,828). After multiple testing correction, the LabWAS of depression PGS revealed significant associations with four elevated immune markers, white blood cell count (WBC) (p-value = 1.07 × 10\textsuperscript{-17}, beta = 0.03, SE = 0.004), urinary WBC (p-value = 1.45 × 10\textsuperscript{-5}, beta = 0.03, SE = 0.007), absolute monocyte count (p-value = 2.54 × 10\textsuperscript{-5}, beta = 0.02, SE = 0.005), and absolute neutrophil count (p-value = 5.91 × 10\textsuperscript{-5}, beta = 0.02, SE = 0.005). Significant associations also comprised several metabolic markers including increased triglycerides (p-value = 3.14 × 10\textsuperscript{-18}, beta = 0.05, SE = 0.006), decreased HDL-C (p-value = 1.23 × 10\textsuperscript{-11}, beta = -0.04, SE = 0.005), decreased calcitriol (p-value = 2.83 × 10\textsuperscript{-8}, beta = -0.04, SE = 0.007), increased glucose (p-value = 2.84 × 10\textsuperscript{-7}, beta = 0.02, SE = 0.004), decreased blood urea nitrogen (BUN) (p-value = 5.19 × 10\textsuperscript{-7}, beta = -0.02, SE = 0.004), decreased calcium (p-value = 9.74 × 10\textsuperscript{-7}, beta = -0.02, SE = 0.004), and decreased calcidiol (p-value = 7.03 × 10\textsuperscript{-5},
beta = -0.04, SE = 0.01). Depression PGS were also associated with decreased troponin-I (p-value = 1.09 × 10^{-6}, beta = -0.05, SE = 0.009), decreased urinary red blood cells (p-value = 1.37 × 10^{-5}, beta = -0.03, SE = 0.006), decreased thyroxine (p-value = 1.72 × 10^{-5}, beta = -0.03, SE = 0.006), and decreased blood carbon dioxide (p-value = 4.06 × 10^{-6}, beta = -0.02, SE = 0.003) (Fig. 1a, Supplementary Table 1).

When controlling for depression diagnosis, depression PGS remained associated with WBC (p-value = 2.54 × 10^{-14}, beta = 0.03, SE = 0.004), triglycerides (p-value = 2.30 × 10^{-11}, beta = 0.05, SE = 0.007), HDL-C (p-value = 9.96 × 10^{-9}, beta = -0.04, SE = 0.007), BUN (p-value = 4.35 × 10^{-6}, beta = -0.02, SE = 0.005), glucose (p-value = 1.43 × 10^{-5}, beta = 0.02, SE = 0.004), calcitriol (p-value = 3.02 × 10^{-5}, beta = -0.04, SE = 0.009), and calcium (p-value = 1.53 × 10^{-4}, beta = 0.02, SE = 0.004) (Supplementary Fig. 1a). After controlling for both depression and anxiety diagnoses, the associations with WBC (p-value = 1.88 × 10^{-15}, beta = 0.03, SE = 0.004), triglycerides (p-value = 7.36 × 10^{-9}, beta = 0.04, SE = 0.008), HDL-C (p-value = 4.54 × 10^{-8}, beta = -0.04, SE = 0.007), glucose (p-value = 1.10 × 10^{-5}, beta = 0.02, SE = 0.005), calcitriol (p-value = 4.69 × 10^{-5}, beta = -0.04, SE = 0.009), and BUN (p-value = 5.70 × 10^{-5}, beta = -0.02, SE = 0.005) survived Bonferroni correction (Supplementary Fig. 1b). In a final sensitivity analysis, we controlled for diagnoses for depression, anxiety, and adjustment reaction as well as median BMI across the EHR. The most significant association remained WBC (p-value = 1.99 × 10^{-13}, beta = 0.03, SE = 0.005), followed by triglycerides (p-value = 7.25 × 10^{-6}, beta = 0.04, SE = 0.008), HDL-C (p-value = 4.77 × 10^{-5}, beta = -0.03, SE = 0.007), and BUN (p-value = 7.23 × 10^{-5}, beta = -0.02, SE = 0.005) (Fig. 1b).

While depression PGS remained robustly associated with WBC across all analyses, the magnitude of the effect was modest (beta = 0.03). Stratification of individuals in the discovery cohort (VUMC) showed that even at the highest decile of depression PGS, WBC measurements are elevated but remain within the clinical reference range (4–11 thousand cells/uL) (Supplementary Fig. 2).

Next, we explored the robustness and generalizability of our findings across ancestry by repeating the depression PGS-LabWAS in individuals of African ancestry in VUMC (N = 12,384). As expected given the reduction in sample size and the poor transferability of PGS to individuals of African ancestry, no lab values were significantly associated with the depression PGS in the African ancestry group. However, the association between depression PGC and WBC trended in the same direction as observed in the European population (p-value = 0.058, beta = 0.02), (Supplementary Table 2).

**Conditional Analyses of WBC**

Depression diagnosis is associated with increased risk for autoimmune disease \(^5\) and is often comorbid with other phenotypes linked to elevated WBC, such as cardiovascular disease \(^2\)–\(^4\), smoking \(^41\),\(^42\), and diabetes \(^6\)–\(^9\). Therefore, we performed a series of conditional analyses to rule out the possibility that a common phenotype associated with both depression PGS and WBC was confounding the observed association. To identify phenotypes associated with both depression PGS and median WBC in VUMC, we performed separate phenome-wide association scans (PheWAS) \(^43\) of depression PGS and WBC measurements. In a PheWAS, ICD codes are hierarchically grouped together into ‘phecodes’ based on
phenotypic similarity. Depression PGS and median WBC were significantly associated with 66 and 469 phecodes, respectively. Of these significantly associated phecodes, 32 were common to both depression PGS and median WBC. Because many of the associated phecodes represented clinically similar conditions, they were binned into seven phe-groups: cardiovascular, psychiatric, obesity, respiratory, hepatic, pain, and autoimmune. (Fig. 2a, Supplementary Table 3).

To determine if comorbidity within a phe-group could account for the relationship between depression PGS and WBC, we controlled for each possible confounding phe-group separately and all phe-groups together. The association between depression PGS and WBC remained significant after controlling for each group separately and controlling for all phenotype groups together (p-value = 4.19 × 10^{-3}, beta = 0.02) with effect estimates similar to the original association (Fig. 2b, Supplementary Table 4).

**Replication in the PsycheMERGE Network**

Given the robustness of the association with WBC and the history of associations between depression status and pro-inflammatory markers, we focused on WBC for replication and further investigation. Replication analysis included fitting the same linear regression model to test the association between depression PGS and median WBC after adjusting for sex, age at median WBC, and genetic ancestry. Findings were replicated in three external biobanks, the Million Veterans Program (MVP), Mount Sinai Icahn School of Medicine (MSSM), and Massachusetts General Brigham Biobank (MGBB) (Supplementary Table 5). In both MVP (N = 289,880) and MGBB (N = 20,828), the association between depression PGS and WBC remained significant with effect estimates similar to those observed at VUMC, even after controlling for depression diagnosis and after controlling for depression and anxiety diagnoses (Fig. 3). In MSSM, the effect size point estimates were similar to those observed in the three other sites, though they did not reach statistical significance, likely due to the smaller sample size (n = 823). The meta-analyzed effect estimate from the four sites was robustly and highly significant (p-value = 2.52 × 10^{-137}, beta = 0.04), even after controlling for depression diagnosis (p-value = 3.93 × 10^{-102}, beta = 0.04), and after controlling for depression and anxiety diagnoses (p-value = 2.48 × 10^{-100}, beta = 0.04) (Fig. 3, Supplementary Table 6).

**Mediation Analysis**

Mediation analyses can help explain and quantify the mechanisms by which an exposure affects an outcome through a mediating variable. Two potential pathways between depression PGS, WBC, and depression diagnosis were assessed using mediation analyses. In the first analysis, median WBC was modeled as a mediator of the relationship between depression PGS (exposure) and depression diagnosis (outcome). Meta-analysis across all sites revealed that WBC mediated 2.0% of the association between depression PGS and depression diagnosis (p-value = 5.05 × 10^{-43}) (Table 1, Supplementary Table 7).
WBC and depression diagnosis mediation analysis. First, WBC was investigated as a mediator between depression PGS and depression diagnosis defined as phecode 296.2. Next, depression diagnosis was assessed as a mediator between depression PGS and WBC. The proportion mediated was calculated by comparing the 50% percentile of depression PGS to the 90th percentile.

| Exposure | Mediator | Outcome | Site | P-value | Proportion Mediated (SE) | 95% Confidence Interval |
|----------|----------|---------|------|---------|--------------------------|-------------------------|
| MDD PGS  | WBC      | MDD diagnosis | MGB  | 0.25    | 0.003 (0.004)            | -0.003–0.011            |
|          |          |          |      |         |                          |                         |
|          | MVP      |          |      | < 2e-16 | 0.035 (0.002)            | 0.031–0.039             |
|          | MSSM     |          |      | 0.871   | -0.016 (0.062)           | -0.255–0.106            |
|          | VUMC     |          |      | 0.16    | 0.003 (0.002)            | -0.001–0.0008           |
|          |          |          |      | 5.05E-43 | 0.02 (0.001)            | 0.017–0.022             |
| MDD PGS  | MDD diagnosis | WBC       | MGB  | 0.24    | 0.018 (0.017)            | -0.014–0.05             |
|          | MVP      |          |      | < 2e-16 | 0.169 (0.009)            | 0.136–0.186             |
|          | MSSM     |          |      | 0.74    | -0.006 (0.674)           | -1.32–1.31              |
|          | VUMC     |          |      | 0.18    | 0.003 (0.002)            | 0–0.007                 |
|          |          |          |      | 2.78E-10 | 0.013 (0.002)           | 0.009–0.017             |

In the second analysis, depression diagnosis was modeled as a mediator of the association between the depression PGS (exposure) and median WBC (outcome). Again, meta-analysis across all sites indicates that depression diagnosis mediated 1.3% of the association between depression PGS and WBC (p-value = 2.78 × 10^{-10}) (Table 1, Supplementary Table 8).

**Depression PGS and WBC-differential Mediation Analysis**

WBC counts are calculated from the sum of five different cell subtypes: neutrophils, lymphocytes, monocytes, basophils, and eosinophils. These cell subtypes can be measured along with the total WBC using a complete blood count differential (CBC-diff) lab. To determine whether specific WBC components accounted for the relationships between depression PGS and total WBC, and between depression PGS and depression diagnosis, we performed a series of multiple mediator analyses. In a multiple mediator analysis, a single main mediator and additional alternative mediators are specified. A structural equation modeling approach is used to assess the effect of the main mediator between the exposure and outcome after controlling for the correlation structure between the alternative mediators and the outcome. Each cell subtype was analyzed as the main mediator with the remaining cell subtypes as the alternative...
mediators. Only differential lab values recorded on the same date (N = 24,383) were used to ensure accurate modeling of the CBC-differential.

When treating depression PGS as the exposure and total WBC as the outcome, neutrophils mediated 69.4% (95% CI: 61.2–73.3%), lymphocytes mediated 22.3% (95% CI: 17.0–24.9%) and monocytes mediated 9.0% (95% CI: 8.2–9.3%) of the association between depression PGS and WBC count. Basophils and eosinophils did not significantly contribute to the correlation between depression PGS and WBC. In the analysis with depression PGS as the exposure and depression diagnosis as the outcome, neutrophils were the only cell type that explained a significant proportion (1.9%; 95% CI: 0.2–3.1%) of the association between depression PGS and depression diagnosis (Table 2).

**Table 2**

| Outcome          | Cell Type  | Proportion Mediated | Lower 95% CI | Upper 95% CI |
|------------------|------------|---------------------|--------------|--------------|
| WBC              | Basophils  | 0.007               | -0.002       | 0.012        |
|                  | Eosinophils| 0.002               | -0.014       | 0.011        |
|                  | Lymphocytes| 0.223               | 0.17         | 0.249        |
|                  | Monocytes  | 0.09                | 0.082        | 0.093        |
|                  | Neutrophils| 0.694               | 0.612        | 0.733        |
| MDD diagnosis    | Basophils  | 0.005               | -0.009       | 0.015        |
|                  | Eosinophils| 0.002               | -0.015       | 0.014        |
|                  | Lymphocytes| 0.008               | -0.007       | 0.018        |
|                  | Monocytes  | -0.001              | -0.018       | 0.011        |
|                  | Neutrophils| 0.019               | 0.002        | 0.031        |

**Discussion**

Depression is consistently associated with increased pro-inflammatory biomarkers, however, the mechanisms underlying these associations remains unclear. In this study, analysis of EHR-linked biobanks within the PsycheMERGE Network were utilized to examine the effect of depression polygenic scores on a variety of clinical lab traits, revealing a robustly replicated association with increased white blood cell count. Notably, several other lab traits associated with depression polygenic scores, including lipids, blood glucose, and blood urea nitrogen. In this study, we chose to further investigate the
relationship with white blood cell count given the robustness of the association to clinical confounders and the existing literature.

In a lab-wide screen, increased polygenic depression risk was associated with increased inflammatory markers including white blood cell count, which was significantly associated with depression PGS even after controlling for depression, anxiety, and multiple comorbid phenotypes, thus highlighting depression polygenic score as an important risk factor for the pro-inflammatory state observed in depression. These results indicate that genetic risk for depression, independent of depression diagnosis or state, are linked to a pro-inflammatory biomarker. The effect of the depression polygenic score on WBC was modest across all biobanks, suggesting that individuals with high depression genetic liability may have an activated, but not abnormal immune system. Nonetheless, sustained activation of the immune system could have important implications for the risk of developing depression.

There are two main models that connect depression to a pro-inflammatory state, the neuroinflammation model and the stress response model. The neuroinflammation model hypothesizes that an activated immune system contributes to risk of developing depression\(^{47,48}\). Alternatively, the stress response model proposes that the stress of depression symptoms leads to a pro-inflammatory state\(^{49,50}\). Importantly, these two models are not mutually exclusive and some have suggested that they form a feedback loop\(^{51,52}\). In support of this hypothesis, the mediation results from this study do not distinguish either the neuroinflammation model or the stress response model as the exclusive pathway between depression and WBC. Rather, the results raise the hypothesis of a “feedback loop” whereby increased depression genetic liability leads to increased WBC, further increasing the likelihood of a depressive episode and subsequent diagnosis. Simultaneously, increased depression genetic liability increases the risk of a depression diagnosis through other mechanisms which in turn, increases WBC, completing the proposed loop.

In the clinic, WBC measurements can be broken down into measurements of each WBC subtype: neutrophils, lymphocytes, monocytes, basophils, and eosinophils. Abnormal levels of different WBC subtypes can index different immune processes. Understanding which cell types underlie the relationship between depression polygenic score and WBC can help narrow a specific immune process that may be altered by depression. In analyses at VUMC, neutrophil counts, lymphocyte counts, and monocyte counts explained the entirety of the association between depression genetics and WBC, while neutrophil counts explained 1.9% of the association between depression polygenic score and depression diagnosis. However, the analysis of WBC subtypes from EHRs comes with limitations: individuals with clinical orders for WBC differential panels may represent a different subsample than those with only the total WBC measurement. Additionally, WBC subtypes contribute unequally to the total WBC measurement. The subtypes that explain the highest percentage of the relationship between depression genetics and WBC mirror the percent contribution to the whole WBC measurement, indicating the analysis might be influenced by cell type proportions even though the differential analyses used cell type count, not proportion.
Overall, using polygenic risk of depression, we prioritized elevated WBC counts as a potential biomarker of depression risk worthy of further investigation. However, our results should be interpreted in light of several analytic limitations. First, the WBC measurements used in the study were clinically derived, with measurements reflecting a range of health states. Because the relationship between depression PGS and WBC remained constant even after controlling for several comorbid phenotypes, it is less likely that the association is due to ascertainment bias for a specific phenotype for which WBC is commonly measured. Additionally, EHRs often contain multiple WBC measurements for the same individual. In this study, only the median values per individual are utilized, leaving unanswered questions about the effect of depression PGS on WBC over time and in response to antidepressant treatment. Finally, even though the relationship between depression PGS and WBC was statistically robust, the effect sizes are small, making WBC an unlikely candidate for use as a diagnostic biomarker of depression. Instead, the associations described in this study highlight the importance of WBC biology in depression and demonstrate the use of EHR-based genomics as a tool for discovery of physiological markers in psychiatric traits.

Online Methods

PsycheMERGE Study Samples

Vanderbilt University Medical Center

Vanderbilt University Medical Center (VUMC) is a tertiary care center that provides inpatient and outpatient care in Nashville, TN. The VUMC electronic health record (EHR) system was established in 1990 and includes data on billing codes from the International Classification of Diseases, 9th and 10th editions (ICD-9 and ICD-10), Current Procedural Terminology (CPT) codes, laboratory values, reports, and clinical documentation. In 2007, VUMC launched a biobank, BioVU, which links a patient’s DNA sample to their EHR. The BioVU consent form is provided to patients in the outpatient 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 (IRB#172020) In VUMC, primary analyses were conducted in individuals of European ancestry, with a sensitivity analysis in individuals of African ancestry.

Million Veteran Program

The Million Veteran Program is an observational cohort study and mega-biobank in the Department of Veterans Affairs (VA)53. Participants are active users of the Veterans Health Administration and provide a blood sample, responses to questionnaires and consent to allow access to clinical data from the VA electronic health records53. The MVP v3.0 data release used in this study includes genotyping data from 455,789 individuals; DNA was extracted from whole blood (which was collected during enrollment to the MVP) and genotyping was performed with the MVP 1.0 Genotyping array53. For this study, we only considered samples with a European Ancestry (EUR) as determined by HARE (Harmonized ancestry and race/ethnicity) analysis54.

Icahn School of Medicine at Mount Sinai
The BioMe Biobank, at the Icahn School of Medicine at Mount Sinai, is an EHR-linked biobank of participants from the Mount Sinai Health System in New York, NY. Participant recruitment into BioMe has been ongoing since 2007, predominantly recruited from general medicine and primary care clinics, and the rest from specialty practices and recruitment events. BioMe participants consent to provide DNA and plasma samples linked to their de-identified EHRs, and then provide additional information on self-reported ancestry, health behaviors, and medical history through questionnaires administered upon enrollment.

Massachusetts General Brigham Biobank (MGBB)

The Massachusetts General Brigham Biobank, formerly known as the Partners Healthcare Biobank, is an ongoing virtual cohort study of patients across the MGB General Brigham hospital system (including Brigham and Women's Hospital, Massachusetts General Hospital, and other affiliated hospitals), which provides a large-scale resource of linked longitudinal electronic health records (EHR) data, genomic data, and self-reported survey data. All patients provided informed consent before enrollment, and all study procedures were approved by the Massachusetts General Brigham Institutional Review Board.

Depression Polygenic Scoring

Depression polygenic scores were generated using PRS-CS using SNP weights from the largest available depression meta-analysis. PRS-CS uses a Bayesian framework to model linkage disequilibrium (LD) from an external reference panel and applies a continuous shrinkage prior on SNP effect sizes. Using PRS-CS-auto, the shrinkage parameter was automatically learnt from the data. The LD reference panel used was constructed from 503 European samples in the 1000 Genomes Project phase 3. For analysis, PGS were scaled to have a mean of zero and a unit standard deviation (SD) so that effect estimates in subsequent analyses are per 1 SD increase in depression PGS.

Lab Quality Control

At VUMC, lab results were extracted from the EHRs of 70,704 individuals and cleaned as previously described. Briefly, labs were required to have at least 70% of observations in a single set of units and filtered for at least 1,000 observations over at least 100 individuals. Observations outside 4 standard deviations of the sample mean were excluded to remove extreme values including those that are biologically implausible. The median observation for each individual in each lab was selected and adjusted for cubic splines of age at measurement. The age-adjusted value was normalized using a rank-based inverse normal transformation (INT) to ensure a normal distribution for downstream analyses, generating age-adjusted INT lab values. For genetic analysis, labs exhibiting no measurable heritability through the GREML analysis in the GCTA software were excluded, leaving 315 labs for use.

LabWAS of Depression PGS in VUMC

Associations between the depression PGS and labs were estimated with a lab-wide association scan (LabWAS) approach in the VUMC sample. LabWAS uses the median, age-adjusted, rank-based inverse normal transformed (INT) lab values from the QualityLab pipeline in a linear regression to determine the association with an input variable after adjusting for covariates. The LabWAS of depression PGS was
controlled for sex and top 10 genetic principal components. Primary analyses were restricted to individuals of European descent. We repeated the depression PGS-LabWAS in individuals of African ancestry in BioVU only (n = 12,384) (Supplementary Fig. 3, Supplementary Table 2). Effect estimates represent per standard deviation increases in depression PGS. In this analysis, the scalability of labWAS across hundreds of labs took precedence over the interpretability of the effect estimate which cannot be converted to the lab's original units.

In conditional analyses, the LabWAS of depression PGS was covaried for median BMI across the EHR, and for depression, anxiety, and adjustment reaction diagnoses, defined by phecodes 296.2, 300.1 and 304, respectively. For all analyses involving phecodes, cases were required to have at least two instances of component ICD codes and controls were required to have zero component ICD codes and zero phecode exclusion codes defined by the phecode map in the PheWAS R package v0.99. Individuals with only one component ICD code were excluded.

Sensitivity Analyses with WBC in VUMC

A series of steps were taken to ensure the association between depression PGS and WBC was not due to a common comorbid confounder phenotype present in individuals with both increased depression PGS and WBC. To find phenotypes associated with both depression PGS and WBC, separate phenome-wide association scans (PheWAS) were conducted of depression PGS and of the median, age-adjusted, INT normalized WBC measurement. Phecode cases were required to have at least two instances of component ICD codes and controls were required to have zero component ICD codes and zero phecode exclusion codes defined by the phecode map in the PheWAS R package v0.99. Individuals with only one component ICD code were excluded. Phecodes were required to have at least 100 cases to be included in the scans. Next, phenotypes that were significantly associated with both depression PGS and WBC at Bonferroni significance (p_{WBC} < 3.64 \times 10^{-5}, p_{depression PGS} < 3.72 \times 10^{-5}) were selected. The common phenotypes were binned into seven categories based on similarity of phecodes: cardiovascular, psychiatric, obesity, respiratory, hepatic, pain, and autoimmune. Group-based case-control variables were constructed, in which an individual was considered a case if they were a case for any of the group's component codes. Controls were required to be a control for all of the component codes. To assess the effect of the comorbid phenotypes on the association between depression PGS and WBC, a series of linear regression analyses were conducted controlling for each of the groups separately. Finally, a linear regression was conducted between depression PGS and WBC controlling for all common phenotype groups. All analyses were controlled for sex and top 10 genetic principal components.

Depression PGS and WBC in the PsycheMERGE Network

Targeted replication analyses between depression PGS and WBC were conducted in three external biobanks. Depression PGS were constructed as described above. In each healthcare system, the median WBC measurement per individual was extracted, age-adjusted for cubic splines of age at measurement, and normalized using a rank-based inverse normal transformation (INT). The depression PGS and WBC were then fitted in a linear regression model controlling for sex and top 10 genetic principal components. The associations controlling for depression and anxiety diagnoses were also replicated using the same
phenotype definition as described in the discovery LabWAS at VUMC. We then meta-analyzed the effect estimates from each analysis across all four sites using a fixed-effect inverse variance weighted model in the meta R package.

**Controlling for the impact of WBC genetics**

We first tested the correlation between WBC PGS and depression diagnosis. The WBC PGS were constructed using PRS-CS-auto with the 1000 Genomes Project Phase 3 European subset reference panel and weights from the UK Biobank WBC summary statistics. WBC PGS were z-score scaled so that the effect estimate is per standard deviation increase in PGS. We first confirmed that the WBC PGS was significantly associated with measured WBC ($p < 2.23 \times 10^{-308}$, beta = 0.14). Then we tested the association between WBC PGS and depression diagnosis (defined as phecode 296.2) across all biobanks, controlling for sex, median age across the medical record, and top 10 genetic principal components using a linear regression model.

In a separate series of analyses, the influence of WBC genetic factors within the depression PGS was examined. First, the effect of genetic regulation of WBC was adjusted from the depression summary statistics by conditioning the depression summary statistics on the WBC summary statistics using multi-trait-based conditional & joint analysis (mtCOJO) in GCTA version 1.91.4. Next, conditioned depression PGS were constructed using the conditioned depression summary statistics. Finally, we tested the association between the conditioned depression PGS and the median age-adjusted INT normalized WBC measurements, controlling for sex and top 10 principal components of ancestry. The effect estimates of each analysis from all four sites were meta-analyzed using a fixed-effect inverse variance weighted model in the meta R package. Results are presented in the Supplementary Results, Supplementary Fig. 4, and Supplementary Table 9.

**Depression PGS and WBC Mediation Analysis**

Two mediation models between depression PGS, WBC, and depression diagnosis were investigated. First, WBC was modeled as the mediator between depression PGS (exposure) and depression diagnosis (outcome). Second, depression diagnosis was modeled as the mediator between depression PGS (exposure) and WBC (outcome). For each analysis, proportion mediated estimates were calculated by dividing individuals in the control and treatment groups based on percentile of depression PGS. The control group consisted of individuals in the 50th percentile of depression PGS. Three different treatment groups were tested: 85th percentile, 90th percentile and 95th percentile. In the main text, results from the 90th percentile treatment group are reported. All results are reported in Supplementary Tables 8 and 9. All analyses were controlled for sex and top 10 genetic principal components. The mediation analyses were conducted using 10,000 bootstrapping simulations until a stable p-value was achieved using the mediation R package. The proportion mediated estimates from all four sites were meta-analyzed using a fixed-effect inverse variance weighted model in the meta R package.

**Depression PGS and WBC-differential Mediation Analysis**
To determine which WBC cell types contributed to the association between depression PGS and WBC measurement or depression diagnosis, a series of multiple mediator analyses were conducted using the mediation\textsuperscript{46} R package. All measurements were required to be recorded on the same date to ensure they were from the same WBC-differential. For individuals with multiple WBC-differentials recorded in their EHR, median WBC values and the corresponding subtype absolute values were selected. All measurements were adjusted for cubic splines of age at observation and normalized\textsuperscript{64,65}.

First, each WBC subtype (basophils, eosinophils, lymphocytes, monocytes, and neutrophils) was analyzed as the main mediator between depression PGS (exposure) and WBC measurement (outcome) with the remaining subtypes as the alternative mediators. Next, each WBC subtype was analyzed as the main mediator between depression PGS (exposure) and depression diagnosis (outcome) with the remaining subtypes as the alternative mediators.

References

1. James, S. L. et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990\textendash2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* \textbf{392}, 1789--1858 (2018).
2. Brown, D. W., Giles, W. H. & Croft, J. B. White blood cell count: An independent predictor of coronary heart disease mortality among a national cohort. *J. Clin. Epidemiol.* \textbf{54}, 316--322 (2001).
3. Hare, D. L., Toukhsati, S. R., Johansson, P. & Jaarsma, T. Depression and cardiovascular disease: A clinical review. *Eur. Heart J.* \textbf{35}, 1365--1372 (2014).
4. Musselman, D. L., Evans, D. L. & Nemeroff, C. B. The relationship of depression to cardiovascular disease: Epidemiology, biology, and treatment. *Arch. Gen. Psychiatry* \textbf{55}, 580--592 (1998).
5. Benros, M. E. et al. Autoimmune diseases and severe infections as risk factors for mood disorders a nationwide study. *JAMA Psychiatry* \textbf{70}, 812--820 (2013).
6. Anderson, R., Freedland, K., RE, C. & Lustman, P. J. The Prevalence of Comorbid Depression. *Diabetes Care* \textbf{24}, 1069--1078 (2001).
7. Gkrania-Klotsas, E. et al. Differential white blood cell count and type 2 diabetes: Systematic review and meta-analysis of cross-sectional and prospective studies. *PLoS One* \textbf{5}, (2010).
8. Mezuk, B., Eaton, W. W., Albrecht, S. & Golden, S. H. Depression and type 2 diabetes over the lifespan: A meta-analysis. *Diabetes Care* \textbf{31}, 2383--2390 (2008).
9. Shim, W. S. et al. The association of total and differential white blood cell count with metabolic syndrome in type 2 diabetic patients. *Diabetes Res. Clin. Pract.* \textbf{73}, 284--291 (2006).
10. Strimbu, K. & Tavel, J. A. What are biomarkers? *Curr. Opin. HIV AIDS* \textbf{5}, 463--466 (2010).
11. Dowlati, Y. et al. A Meta-Analysis of Cytokines in Major Depression. *Biol. Psychiatry* \textbf{67}, 446--457 (2010).
12. Dinan, T. G. Inflammatory markers in depression. *Curr. Opin. Psychiatry* \textbf{22}, (2009).
13. Howren, M. B., Lamkin, D. M. & Suls, J. Associations of Depression With C-Reactive Protein, IL-1, and IL-6: A Meta-Analysis. *Psychosom. Med.* **71**, (2009).

14. Euteneuer, F. *et al.* Peripheral immune alterations in major depression: The role of subtypes and pathogenetic characteristics. *Front. Psychiatry* **8**, 8–13 (2017).

15. Ancelin, M.-L. *et al.* Gender and Genotype Modulation of the Association Between Lipid Levels and Depressive Symptomatology in Community-Dwelling Elderly (The ESPRIT Study). *Biol. Psychiatry* **68**, 125–132 (2010).

16. Gupta, A., Petkar, S. B., Jadhav, A. A. & Dubey, V. Study of Lipid Derangement in Pyschiatric Disorder. *Ind. Med. Gaz.* 253–256 (2013).

17. Partonen, T., Haukka, J., Virtamo, J., Taylor, P. R. & Lönnqvist, J. Association of low serum total cholesterol with major depression and suicide. *Br. J. Psychiatry* **175**, 259–262 (1999).

18. Evans, S. J. *et al.* Dysregulation of the fibroblast growth factor system in major depression. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15506–15511 (2004).

19. Iga, J. *et al.* Gene expression and association analysis of vascular endothelial growth factor in major depressive disorder. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* **31**, 658–663 (2007).

20. Kahl, K. G. *et al.* Angiogenic factors in patients with current major depressive disorder comorbid with borderline personality disorder. *Psychoneuroendocrinology* **34**, 353–357 (2009).

21. Aydemir, C. *et al.* Brain-derived neurotrophic factor (BDNF) changes in the serum of depressed women. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* **30**, 1256–1260 (2006).

22. Molendijk, M. L. *et al.* Serum levels of brain-derived neurotrophic factor in major depressive disorder: state–trait issues, clinical features and pharmacological treatment. *Mol. Psychiatry* **16**, 1088–1095 (2011).

23. Dwivedi, Y. Brain-derived neurotrophic factor: role in depression and suicide. *Neuropsychiatr. Dis. Treat.* **5**, 433–449 (2009).

24. Castrén, E. & Rantamäki, T. The role of BDNF and its receptors in depression and antidepressant drug action: Reactivation of developmental plasticity. *Dev. Neurobiol.* **70**, 289–297 (2010).

25. Euesden, J., Danese, A., Lewis, C. M. & Maughan, B. A bidirectional relationship between depression and the autoimmune disorders - New perspectives from the National Child Development Study. *PLoS One* **12**, 1–14 (2017).

26. Ramírez, K., Fornaguera-Trías, J. & Sheridan, J. F. Stress-Induced Microglia Activation and Monocyte Trafficking to the Brain Underlie the Development of Anxiety and Depression BT - Inflammation-Associated Depression: Evidence, Mechanisms and Implications. in (eds. Dantzer, R. & Capuron, L.) 155–172 (Springer International Publishing, 2017). doi:10.1007/7854_2016_25

27. Hasselmann, H. *et al.* Pro-inflammatory Monocyte Phenotype and Cell-Specific Steroid Signaling Alterations in Unmedicated Patients With Major Depressive Disorder. *Front. Immunol.* **9**, 2693 (2018).

28. Wohleb, E. S. & Delpech, J.-C. Dynamic cross-talk between microglia and peripheral monocytes underlies stress-induced neuroinflammation and behavioral consequences. *Prog. Neuro-*
29. Seidel, A. et al. Major depressive disorder is associated with elevated monocyte counts. *Acta Psychiatr. Scand.* **94**, 198–204 (1996).

30. Gaarden, T. L. et al. Exploration of 27 plasma immune markers: a cross-sectional comparison of 64 old psychiatric inpatients having unipolar major depression and 18 non-depressed old persons. *BMC Geriatr.* **18**, 149 (2018).

31. Lynall, M.-E. et al. Peripheral Blood Cell &x2013;Stratified Subgroups of Inflamed Depression. *Biol. Psychiatry* (2020). doi:10.1016/j.biopsych.2019.11.017

32. Kayhan, F., Gündüz, Ş., Ersoy, S. A., Kandeğer, A. & Annagür, B. B. Relationships of neutrophil-lymphocyte and platelet-lymphocyte ratios with the severity of major depression. *Psychiatry Res.* **247**, 332–335 (2017).

33. Demir, S. et al. Neutrophil-lymphocyte ratio in patients with major depressive disorder undergoing no pharmacological therapy. *Neuropsychiatr. Dis. Treat.* **11**, 2253–2258 (2015).

34. De Berardis, D. et al. The Role of C-Reactive Protein in Mood Disorders. *Int. J. Immunopathol. Pharmacol.* **19**, 721–725 (2006).

35. Wray, N. R. et al. From Basic Science to Clinical Application of Polygenic Risk Scores. *JAMA Psychiatry* (2020). doi:10.1001/jamapsychiatry.2020.3049

36. Estiri, H., Klann, J. G. & Murphy, S. N. A clustering approach for detecting implausible observation values in electronic health records data. *BMC Med. Inform. Decis. Mak.* (2019). doi:10.1186/s12911-019-0852-6

37. Pivovarov, R., Albers, D. J., Sepulveda, J. L. & Elhadad, N. Identifying and mitigating biases in EHR laboratory tests. *J. Biomed. Inform.* (2014). doi:10.1016/j.jbi.2014.03.016

38. Perrotta, P. L. & Karcher, D. S. Validating laboratory results in electronic health records: A college of American pathologists Q-probes study. *Archives of Pathology and Laboratory Medicine* (2016). doi:10.5858/arpa.2015-0320-CP

39. Dennis, J. K. et al. Lab-wide association scan of polygenic scores identifies biomarkers of complex disease. *medRxiv* 2020.01.24.20018713 (2020). doi:10.1101/2020.01.24.20018713

40. Zheutlin, A. et al. VALIDATION OF PSYCHIATRIC POLYGENIC RISK SCORES ACROSS THREE HEALTHCARE SYSTEMS USING ELECTRONIC HEALTH RECORDS. *Eur. Neuropsychopharmacol.* (2019). doi:10.1016/j.euroneuro.2018.07.068

41. Glassman, A. H. et al. Smoking, Smoking Cessation, and Major Depression. *JAMA J. Am. Med. Assoc.* **264**, 1546–1549 (1990).

42. Sunyer, J. et al. Longitudinal relation between smoking and white blood cells. *Am. J. Epidemiol.* **144**, 734–741 (1996).

43. Denny, J. C. et al. PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btq126
44. Keller, M. F. et al. Trans-ethnic meta-analysis of white blood cell phenotypes. *Hum. Mol. Genet.* **23**, 6944–6960 (2014).

45. MacKinnon, D. P., Fairchild, A. J. & Fritz, M. S. Mediation Analysis. *Annu. Rev. Psychol.* **58**, 593–614 (2007).

46. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. Mediation: R package for causal mediation analysis. *J. Stat. Softw.* **59**, 1–38 (2014).

47. Troubat, R. et al. Neuroinflammation and depression: A review. *Eur. J. Neurosci.* 1–21 (2020).

doi:10.1111/ejn.14720

48. Raison, C. L., Capuron, L. & Miller, A. H. Cytokines Sing the Blues: Inflammation and the Pathogenesis of Depression. *Trends Immunol.* **27**, 24–31 (2006).

49. Olff, M. Stress, depression and immunity: the role of defense and coping styles. *Psychiatry Res.* **85**, 7–15 (1999).

50. Bao, A. M. & Swaab, D. F. The human hypothalamus in mood disorders: The HPA axis in the center. *IBRO Reports* **6**, 45–53 (2019).

51. Finnell, J. E. & Wood, S. K. Neuroinflammation at the interface of depression and cardiovascular disease: Evidence from rodent models of social stress. *Neurobiology of Stress* (2016).

doi:10.1016/j.ynstr.2016.04.001

52. Hurley, L. L. & Tizabi, Y. Neuroinflammation, neurodegeneration, and depression. *Neurotoxicity Research* (2013). doi:10.1007/s12640-012-9348-1

53. Gaziano, J. M. et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).

54. Fang, H. et al. Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. *Am. J. Hum. Genet.* **105**, 763–772 (2019).

55. Karlson, E. W., Boutin, N. T., Hoffnagle, A. G. & Allen, N. L. Building the partners healthcare biobank at partners personalized medicine: Informed consent, return of research results, recruitment lessons and operational considerations. *J. Pers. Med.* **6**, 1–11 (2016).

56. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1–10 (2019).

57. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).

58. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

59. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

60. Denny, J. C. et al. PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010).

61. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, (2015).
62. Astle, W. J. et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429.e19 (2016).

63. Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, (2018).

64. Beasley, T. M., Erickson, S. & Allison, D. B. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav. Genet.* (2009). doi:10.1007/s10519-009-9281-0

65. McCaw, Z. R., Lane, J. M., Saxena, R., Redline, S. & Lin, X. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics* (2019). doi:10.1111/biom.13214