nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-019-0759-3

Genomic prediction of depression risk and resilience under stress

Yu Fang[1], Laura Scott[2], Peter Song[2], Margit Burmeister[1] and Srijan Sen[1*]

Advancing ability to predict who is likely to develop depression holds great potential in reducing the disease burden. Here, we use the predictable and large increase in depression with physician training stress to identify predictors of depression. Applying the major depressive disorder polygenic risk score (MDD-PRS) derived from the most recent Psychiatric Genomics Consortium–UK Biobank–23andMe genome-wide association study to 5,227 training physicians, we found that MDD-PRS predicted depression under training stress ($\beta = 0.095$, $P = 4.7 \times 10^{-16}$) and that MDD-PRS was more strongly associated with depression under stress than at baseline (MDD-PRS × stress interaction $\beta = 0.036$, $P = 0.005$). Further, known risk factors accounted for substantially less of the association between MDD-PRS and depression when under stress than at baseline, suggesting that MDD-PRS adds unique predictive power in depression prediction. Finally, we found that low MDD-PRS may have particular use in identifying individuals with high resilience. Together, these findings suggest that MDD-PRS holds promise in furthering our ability to predict vulnerability and resilience under stress.

According to the World Health Organization, depression is the leading cause of disease-associated disability in the world[1]. As current treatments for depression only result in remission in a few cases and new treatments have been slow to emerge, the burden of depression, including suicide, has continued to grow[2,3].

In populations at high risk, prevention of depression may be an effective strategy. The US National Academy of Medicine has highlighted the need to develop, evaluate and implement prevention interventions for depression and other mental, emotional and behavioural disorders[4,5]. However, our current ability to predict those most at risk for depression is limited[6].

Genetic variation accounts for 30–40% of the population variation in unipolar depression risk[7]. Since 2015, genome-wide association studies have, for the first time, identified many variants associated with depression[8,9]. The number of risk variants identified has grown progressively as the number of subjects from the Psychiatric Genomics Consortium (PGC), UK Biobank, 23andMe and other large datasets has increased[10–13]. No individual variants of moderate to large effect have emerged, with evidence indicating that risk for depression is distributed across genome[12,13]. Because the effect sizes of identified depression variants are small, any individual polymorphism has limited use for risk prediction. Polygenic risk scores (PRS) provide a mechanism for aggregating the cumulative impact of common polymorphisms by summing the number of risk variant alleles in each individual weighted by the impact of each allele on risk of disease[14,15]. In other disease phenotypes, PRS has shown value in predicting disease. For instance, the PRS for cardiovascular disease substantially improves risk prediction for disease beyond known risk factors[16].

Prospective cohort studies are critical to evaluating the predictive power of PRS[15]. With life stress accounting for 30–40% of the population variation in unipolar depression risk[17] and since approximately 80% of depressive episodes are preceded by a major stressor[18], a promising strategy is to assess the major depressive disorder PRS (MDD-PRS) as a predictor for the development of depression under stress. However, the unpredictable nature of stress makes prospective studies of depression difficult. The first year of professional physician training, medical internship, is an unusual situation where the onset of stress can be reliably predicted. A meta-analysis of 54 studies and over 17,000 subjects established that the prevalence of depression among training physicians is 28.8%, with a fivefold to sixfold increase with the onset of internship[19]. Work with early cohorts of the Intern Health Study identified neuroticism, history of depression and a difficult early family environment as factors that predict the development of depression during internship stress[20]. Here, we use internship as a model to assess the predictive power of MDD-PRS for depression under stress.

Results

A total 5,227 medical interns of European ancestry from the Intern Health Study were included for analysis: 50.3% of the sample were women and the mean age was 27.6 yr. We measured depressive symptoms using the nine-item patient health questionnaire (PHQ-9) before internship year started (baseline) and every 3 months during the stressful internship year. Included participants completed the baseline survey and at least one quarterly survey (mean number of quarterly survey completions were 3.46, s.d. = 0.90). Raw PHQ-9 scores range from 0 to 27. The participants had a mean PHQ-9 score of 2.5 (s.d. = 2.9) at baseline and 5.6 (s.d. = 3.8) during internship. Of the subjects, 3.4% met PHQ depression criteria (PHQ-9 score ≥ 10) at baseline (Table 1), with 33.2% of subjects meeting PHQ criteria for depression at least once during internship. The average PHQ-9 score and PHQ depression rates were 5.4 (s.d. = 4.2); and 14.9%, 5.6 (s.d. = 4.4), 15.8%, 5.6 (s.d. = 4.5), 15.0%, 5.4 (s.d. = 4.6) and 15.0% at the 3-, 6-, 9- and 12-month assessments, respectively.

Association of MDD-PRS with PHQ-9 depressive symptom score. We used the summary statistics derived from the most recent MDD genome-wide association study (GWAS), a meta-analysis of the PGC MDD phase 2, UK Biobank and 23andMe (a personal genetics company)[13], to calculate the MDD-PRS. The standardized MDD-PRS in intern subjects had a near-normal distribution (Fig. 1).

[1]Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI, USA. [2]School of Public Health, University of Michigan, Ann Arbor, MI, USA. *e-mail: srijan@umich.edu

Table 1 | Intern sample characteristics ($n = 5,227$)

Years age, mean (s.d.)	27.6 (2.7)
Female sex, *n* (%)	2,627 (50.3%)
Personal history of depression, *n* (%)	2,434 (46.6%)
Neuroticism score, mean (s.d.)	21.2 (8.8)
Early family environment score, mean (s.d.)	11.7 (8.8)
Baseline depression symptom score (PHQ-9), mean (s.d.)	2.5 (2.9)
Baseline PHQ depression, *n* (%)	176 (3.4%)

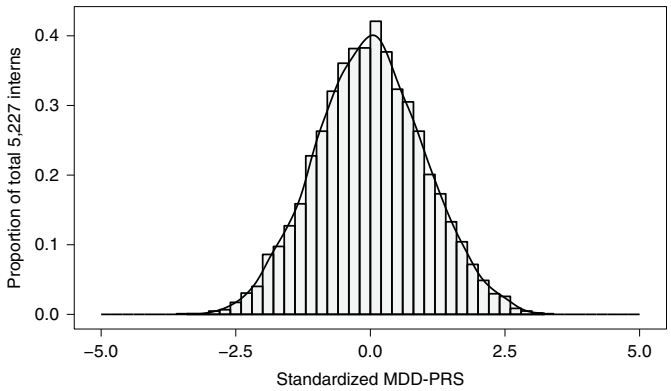


Fig. 1 | MDD-PRS distribution. MDD-PRS has a near-normal distribution in Intern Health Study samples ($n = 5,227$). Represented on the *x* axis, MDD-PRS was mean-centred and scaled to s.d. = 1.

To compare the predictive power of MDD-PRS on depression at baseline and during internship stress, we assessed the association between MDD-PRS and inverse-normalized PHQ-9 score under both conditions. After adjustment for age, sex and the top ten genotype-based principal components (PCs), we found that 1 s.d. increase of MDD-PRS was associated with 0.063 higher inverse-normalized PHQ-9 score at baseline ($P = 5.4 \times 10^{-6}$) and 0.095 higher inverse-normalized PHQ-9 score during internship ($P = 4.7 \times 10^{-16}$) (Table 2 and Fig. 2a,b left diagrams). To assess the interaction between MDD-PRS and stress, we included both the main effect of MDD-PRS and internship stress in a linear mixed model, added the MDD-PRS × internship stress interaction term, and adjusted for the same covariates above, along with the interaction terms for each covariate with MDD-PRS and internship stress. We found a significant interaction between MDD-PRS and internship stress status on PHQ-9 depressive symptom score ($\beta = 0.036$, $P = 0.005$), indicating that the effect of MDD-PRS on PHQ-9 score was greater under internship stress than at baseline (Table 2). To further assess whether the findings were an artefact of latent population stratification, we conducted a sensitivity analysis using 5,116 subjects who fit a more conservative definition of European ancestry (see Methods) and obtained results that were fundamentally unchanged. Similarly, to assess the robustness of findings to the inclusiveness of PRS construction approaches, we conducted another sensitivity analysis with MDD-PRS generated with *P*-value thresholds other than 1 (5×10^{-8}, 1×10^{-5}, 1×10^{-4}, 1×10^{-3}, 0.01, 0.05, 0.1 and 0.5). We found significant associations with both baseline and internship PHQ-9 scores across all *P*-value thresholds (baseline PHQ-9 association, $P = 0.064$ to 8.1×10^{-6}; internship PHQ-9 associations, $P = 0.032$ to 5.9×10^{-16}). With more inclusive thresholds (0.1 and 0.5), the interaction between PRS and stress remained essentially the same ($\beta = 0.028$ and 0.03; $P = 0.024$ and 0.018) but became non-significant at less inclusive thresholds.

Finally, to test the robustness of our findings to the set of variants used to calculate MDD-PRS, we also calculated MDD-PRS using linkage disequilibrium (LD) pruned-imputed common single nucleotide polymorphisms (SNPs). Comparing to genotype data derived MDD-PRS, we observed slightly attenuated but still significant associations between imputed data derived MDD-PRS and PHQ-9, both at baseline ($\beta = 0.063$, $P = 4.8 \times 10^{-6}$) and during internship ($\beta = 0.092$, $P = 3.2 \times 10^{-15}$) and interaction ($\beta = 0.034$, $P = 0.008$) (Extended Data Fig. 1).

In addition to the quantitative PHQ-9 score, we also used PHQ depression diagnosis as an outcome measure. In a logistic regression, we found no significant association between MDD-PRS and depression diagnosis at baseline (odds ratio OR = 1.01, $P = 0.87$). In contrast, MDD-PRS was significantly associated with depression diagnosis during internship (OR = 1.21, $P = 1.9 \times 10^{-10}$). Parallel to the findings with the quantitative PHQ-9 score, there was a significant interaction between MDD-PRS and internship status on PHQ depression diagnosis (OR = 1.40, $P = 0.003$, additive interaction, relative excess risk due to interaction = 31.75, 95% confidence interval CI = 5.47–58.03) indicating the effect of MDD-PRS on depression prevalence was greater during internship than baseline.

Because the MDD GWAS meta-analysis results were generated using European ancestry individuals, we restricted our main MDD-PRS analysis to the European ancestry subjects from our sample[21]. To explore the predictive ability of European ancestry-based MDD-PRS in individuals of other ancestries, we assessed the association between MDD-PRS and PHQ-9 score in Intern Health Study participants of East Asian ancestry ($n = 816$) and South Asian ancestry

Table 2 | MDD-PRS associations with PHQ-9 depressive symptom scores ($n = 5,227$)

Inclusion of known baseline risk factors[a] as covariates	Time point of PHQ-9 score	β (s.e.m.)		*P*	
		MDD-PRS	MDD-PRS × internship stress interaction[b]	MDD-PRS	MDD-PRS × internship stress interaction[b]
Not included	Baseline	0.063 (0.014)		5.38×10^{-6}	
	Internship	0.095 (0.012)		4.68×10^{-16}	
	Baseline and internship	0.108 (0.034)	0.036 (0.013)	0.001	0.005
Included	Baseline	0.012 (0.012)		0.34	
	Internship	0.046 (0.010)		6.87×10^{-6}	
	Baseline and internship	0.078 (0.030)	0.035 (0.013)	0.009	0.006

[a]Neuroticism, personal history of depression, early family environment. [b]MDD-PRS × internship stress interaction term is only available in the model including both baseline and internship data.

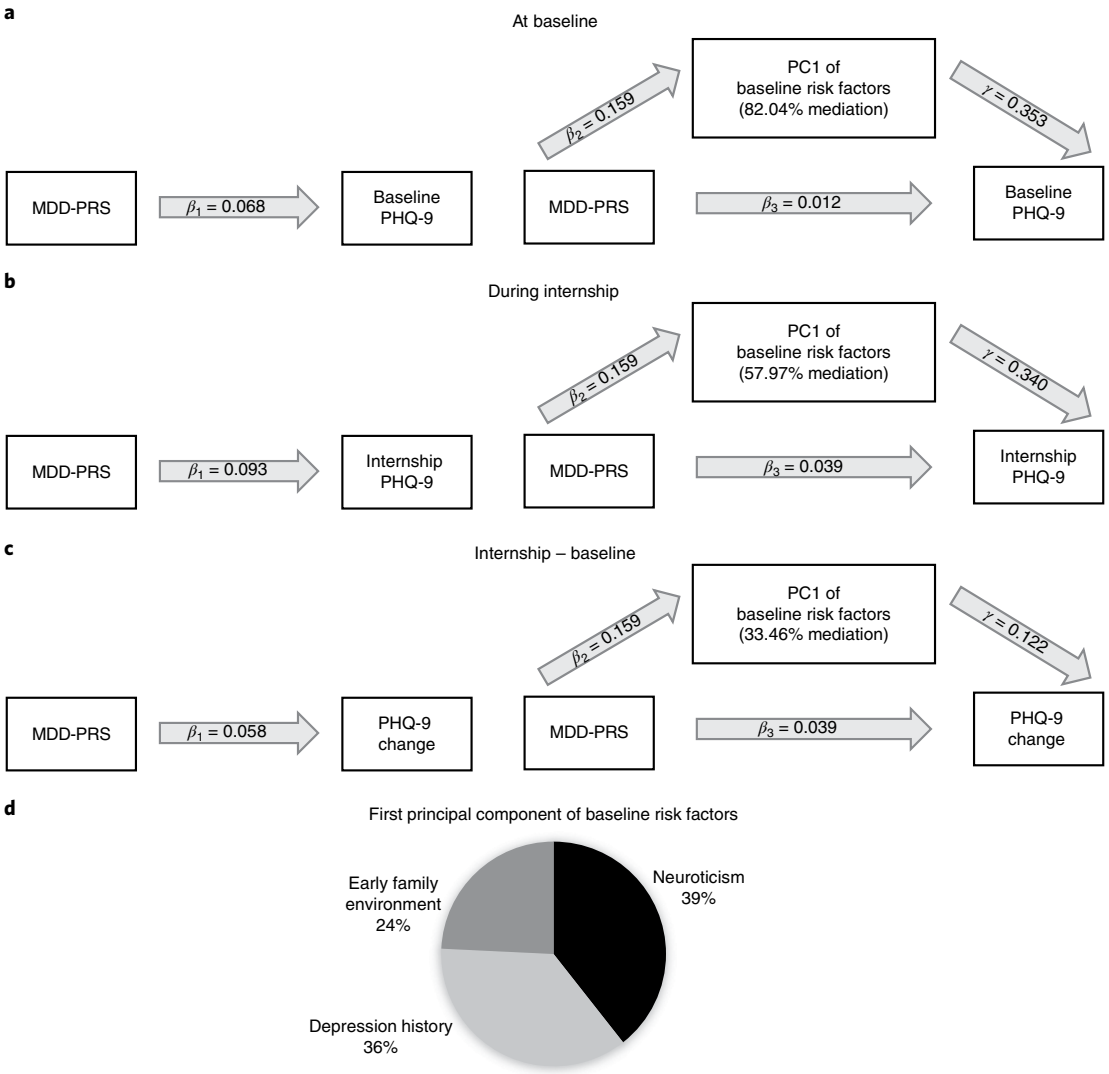


Fig. 2 | Associations of MDD-PRS and PHQ-9 depressive symptom score and mediations of the associations by known risk factors. a–c, The associations of MDD-PRS and PHQ-9 score at baseline, during internship and change, without mediator (left diagram) and mediated by known baseline risk factors (right diagram). In **a**, all $P<2\times10^{-16}$ except for $P(\beta_3)=0.32$. In **b**, all $P<2\times10^{-16}$ except for $P(\beta_3)=1.26\times10^{-4}$. In **c**, all $P<2\times10^{-16}$ except for $P(\beta_3)=0.005$. **d**, Percentage of variance in first PC explained by each risk factor.

($n=595$). We did not find any significant association between MDD-PRS and PHQ-9 scores in either the East Asian group (baseline $\beta=0.021$, $P=0.55$; during internship $\beta=0.036$, $P=0.21$) or the South Asian group (baseline $\beta=0.077$, $P=0.06$; internship $\beta=0.048$, $P=0.18$).

Mediation of the association between MDD-PRS and PHQ-9 depressive symptom score by known risk factors. Neuroticism has been identified as a mediator of the association between depression PRS and self-reported and clinical depression[22]. In this study, we conducted mediation analysis to quantify the proportion of the association mediated by three risk factors—neuroticism, personal history of depression and stressful early family environment—previously demonstrated to predict depression in both the general population and training physicians specifically[20]. To capture the joint contributions of the three known depression risk factors, we performed principal component analysis (PCA) and used the first PC in our analysis (known risk factor-based PC). The contributions of neuroticism, personal history of depression and early family environment to the first PC were 39%, 36% and 24%, respectively (Fig. 2d). In the mediation model, the total effect of MDD-PRS on PHQ-9 consists of the indirect effect working through the known risk factor-based PC and the direct effect of MDD-PRS on PHQ-9 score (equation details in Methods). The known risk factor-based PC explained 82.04% of the association between MDD-PRS and PHQ-9 score at baseline but only 57.97% during internship (Table 3 and Fig. 2a,b right diagrams). As demonstrated in right diagrams in Fig. 2a,b, the indirect effects of the known risk factor-based PC on PHQ-9 score were similar at baseline ($\beta_2\gamma=0.159\times0.353=0.056$) and during internship ($\beta_2\gamma=0.159\times0.340=0.054$). Therefore, we further tested the mediation effect of the known risk factor-based PC on the association of MDD-PRS and PHQ-9 change (internship–baseline) and found that it only explained 33.46% of the association.

Differentiation of high-risk/high-resilience subjects. Khera et al. identified that individuals in the extreme high tail of cardiovascular disease-PRS distribution have higher risk of disease compared to other individuals[16]. To assess if either extreme tail of the PRS dis-

Table 3 | Mediation test of baseline predictors on MDD genetic score predicting PHQ-9 at baseline and during internship

Baseline risk factor	PHQ-9 score at baseline			Average PHQ-9 score during internship		
	Mediation (%)	95% CI	P^b	Mediation (%)	95% CI	P^b
Neuroticism	57.37%	(39.28%, 88.92%)	$<1\times10^{-5}$	38.70%	(28.14%, 50.87%)	$<1\times10^{-5}$
Personal history of depression	41.89%	(28.22%, 68.65%)	$<1\times10^{-5}$	32.42%	(24.08%, 43.78%)	$<1\times10^{-5}$
Early family environment	21.54%	(13.08%, 36.48%)	$<1\times10^{-5}$	14.97%	(9.64%, 21.80%)	$<1\times10^{-5}$
First principal component of three risk factors[a]	82.04%	(58.66%, 98.24%)	$<1\times10^{-5}$	57.97%	(46.13%, 74.02%)	$<1\times10^{-5}$

[a]Neuroticism, personal history of depression and early family environment. [b]*P* value was obtained by 100,000 non-parametric bootstraps.

tribution in our sample more effectively differentiated resilience versus susceptibility to depression during internship stress, we followed the approach of Khera et al. and divided our subjects into 40 quantiles from low to high MDD-PRS, each with 2.5% of the sample ($n=131$). Figure 3 displays baseline and internship PHQ depression proportion of each of the 40 MDD-PRS groups. In addition, Extended Data Fig. 2 displays the baseline and internship average PHQ-9 score of each of the 40 groups.

To quantitatively assess for a difference in depression risk prediction power between the lowest and highest tail, we serially dichotomized the sample using different MDD-PRS percentile cutpoints and compared the proportion of individuals with depression in the subjects above and below each of the cutpoints. Subjects in the lowest 2.5% MDD-PRS distribution (low tail; $n=131$) had lower rates of PHQ depression during internship (12.2%) compared to the remaining sample (33.8%) (OR = 0.27, 95% CI = 0.15–0.46, $P=3.0\times10^{-8}$). In contrast, depression proportion of subjects with MDD-PRS scores in the top 2.5% (high tail; 36.6%) did not differ significantly from depression proportion of the remaining sample (33.1%; OR = 1.17, 95% CI = 0.80–1.69, $P=0.40$). Using a permutation test with 10,000 permutations, we assessed if the OR of the depressed subject for one tail was greater than the OR for the other tail (the reference sample for each OR being the subset of participants with lower MDD-PRS). For the 2.5% cutpoint, we found the low tail had a significantly larger OR than the high tail ($P<1\times10^{-4}$, Table 4). Similarly, using the PHQ-9 depressive symptom score, we found the test statistic for the lowest tail was significantly larger than that for the high tail ($P=8\times10^{-4}$, Table 5). These results indicate the lowest 2.5% of the MDD-PRS distribution better differentiates depression risk and resilience compared to the upper 2.5% of the distribution.

When we tested more inclusive tail cutpoints, we found that the differences in the proportion depression between the low MDD-PRS group and the remaining sample remained significant—for lower tail cutpoints at 5% ($P=2.0\times10^{-5}$), 10% ($P=5.8\times10^{-5}$) and 25% ($P=8.3\times10^{-5}$) (Table 4). The differences in the depression proportion between the high MDD-PRS group and the remaining sample were also significant for upper tail cutpoints of 5% (5.9×10^{-4}), 10% ($P=0.002$) and 25% ($P=1.9\times10^{-5}$) (Table 4). The bottom OR and top OR were not significantly different for any of the more inclusive cutpoints. We saw a similar pattern for PHQ-9 score as the outcome measure (Table 5), indicating no differences from what would be expected by chance for more inclusive cutpoints.

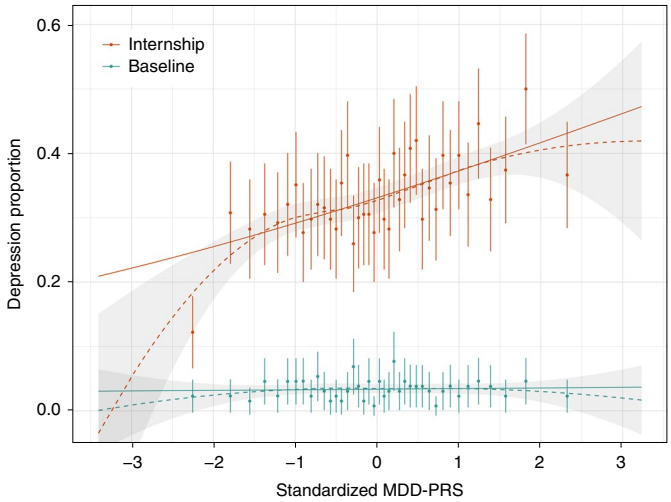


Fig. 3 | PHQ depression proportion by MDD-PRS group. A total 5,227 subjects from Intern Health Study were binned into 40 groups of 2.5% of subjects ($n=131$ per group) from low to high MDD-PRS (left to right). The 40 *x* axis groups are defined by group-wise average standardized MDD-PRS. The proportion of subjects meeting criteria for PHQ depression at baseline (cyan dots) and during internship (orange dots) are plotted with 95% CI error bars. Locally estimated scatterplot smoothing (LOESS) fitting line (dash line) shadowed by 95% CI and logistic regression fitting line (solid line) were applied to both baseline and internship plots. Optimal span parameter for LOESS regression was selected by generalized cross-validation method.

Discussion

Building on the success of recent large-scale GWAS for MDD, this investigation uses a prospective cohort design to demonstrate that MDD-PRS is a significant predictor of future depression. Further, we find evidence that the association between MDD-PRS and depression is stronger in the presence of stress and that the additional predictive power of MDD-PRS under stress is largely independent of known risk factors for depression. Finally, we found that low MDD-PRS scores may have particular use in identifying individuals highly resilient to stress.

Our finding that MDD-PRS associates with depression during internship stress provides empirical evidence that the cumulative impact of common polymorphisms can produce meaningful risk prediction for depression. We found that individuals in the lowest 2.5% of the PRS distribution in our sample were three times less likely to develop depression under stress than were the rest of the sample. The predictive power of MDD-PRS will probably continue to improve with increasingly larger discovery GWAS studies. On the basis of the improving prediction profiles, individuals could be stratified into different strata of risk for transition to depression[23], with intensive prevention strategies targeted to high-risk individuals. For instance, in the population of training physicians, web-based cognitive behavioural therapy has been shown to be effective in the prevention of depression and suicidal ideation and could be targeted for prevention[24].

Table 4 | Comparison of internship PHQ depression in high and low MDD-PRS groups with the remaining sample across PRS cutoffs

PRS tail percentile (%)	Tail (*n*)	Remaining (*n*)	Tail	Depression percentage (tail)	Depression percentage (remaining)	OR (95% CI)	*P*, Fisher's exact test, OR	*P*, test for difference in magnitude of the high and low tail OR[a]
2.5	131	5,096	Low	12.2%	33.8%	0.27 (0.15–0.46)	2.98×10^{-8}	$<1\times10^{-4}$
			High	36.6%	33.1%	1.17 (0.80–1.69)	0.40	
5	262	4,965	Low	21.4%	33.8%	0.53 (0.39–0.72)	2.02×10^{-5}	0.32
			High	43.1%	32.7%	1.56 (1.20–2.02)	5.86×10^{-4}	
10	523	4,704	Low	25.4%	34.1%	0.66 (0.53–0.81)	5.79×10^{-5}	0.33
			High	39.2%	32.5%	1.34 (1.10–1.61)	0.002	
25	1,307	3,920	Low	28.8%	34.7%	0.76 (0.66–0.87)	8.28×10^{-5}	0.86
			High	38.1%	31.6%	1.33 (1.17–1.52)	1.90×10^{-5}	
50	2,613	2,614	Low	29.9%	36.6%	0.74 (0.66–0.83)	2.71×10^{-7}	-

[a]Test statistic = (1/low tail OR)/ high tail OR, evaluated by permutation test (see Methods).

Table 5 | Comparison of average internship PHQ-9 scores in high and low MDD-PRS groups with the remaining sample across PRS cutoffs average internship PHQ-9 scores in MDD-PRS groups

PRS tail percentile (%)	Tail (*n*)	Remaining (*n*)	Tail	PHQ-9 (tail)	PHQ-9 (remaining)	ΔPHQ, tail—remaining (95% CI)	*P*, two sample *t*-test, ΔPHQ	*P*, test of difference between high and low tail ΔPHQ[a]
2.5	131	5,096	Low	3.87	5.62	−1.75 (−2.26 to −1.24)	3.53×10^{-10}	8×10^{-4}
			High	6.34	5.56	0.78 (0.07 to 1.50)	0.033	
5	262	4,965	Low	4.34	5.64	−1.30 (−1.71 to −0.88)	2.88×10^{-9}	0.37
			High	6.82	5.51	1.31 (0.78 to 1.84)	1.73×10^{-6}	
10	523	4,704	Low	4.78	5.67	−0.88 (−1.20 to −0.57)	3.52×10^{-8}	0.28
			High	6.27	5.50	0.77 (0.40 to 1.14)	4.35×10^{-5}	
25	1,307	3,920	Low	5.12	5.73	−0.62 (−0.85 to −0.38)	2.09×10^{-7}	0.70
			High	6.11	5.40	0.71 (0.46 to 0.96)	1.79×10^{-8}	
50	2,613	2,614	Low	5.29	5.86	−0.58 (−0.78 to −0.37)	4.72×10^{-8}	-

[a]Test statistic = ($-t$-statistic $_{\text{low tail PHQ-9 - remaining PHQ-9}}$) − ($t$-statistic $_{\text{high tail PHQ-9 - remaining PHQ-9}}$), evaluated by permutation test (see Methods).

Under baseline, low stress conditions, the link between MDD-PRS and depression is largely explained by established risk factors measured in the study. In contrast, the established risk factors only explained about half the association between MDD-PRS and depression under stress conditions. Understanding the outstanding mechanisms through which genomic predisposition leads to depression under stress could help to better elucidate how stress gets 'under the skin' and exerts pathogenic effects. Further, effective risk predictors for depression will ultimately incorporate genetic variables with other established predictors. The finding that only about half the predictive power of MDD-PRS is mediated by established predictors of depression under internship stress suggests that genomics can add meaningful explanatory power to risk prediction from established factors.

We also find preliminary evidence that the overall association between MDD-PRS and depression under stress is driven disproportionately by the lower end of the PRS distribution. Specifically, while individuals with very low PRS scores are substantially less likely to become depressed relative to the rest of the sample, individuals with very high PRS scores do not show an analogously higher relative risk of depression. As a result, MDD-PRS may be better at identifying resilient individuals than at identifying those that are most at risk for depression under stress. With the relatively small number of subjects ($n=131$) in each of the PRS subgroups, these findings should be assessed in other prospective stress samples before drawing definitive conclusions. Resilience is a dynamic and active neurophysiological and psychological response to stress that is not merely the absence of vulnerability[25]. Delineating the genomic factors that are protective against disease has informed genomically anchored medicines that assist in maintaining health[26]. Identifying specific genes within the broad PRS that are aiding resilience has the potential to point to new therapies for depression.

There are some limitations to this study. First, this study focused on physician training as a specific stressor. While the established predictors of depression are similar to predictors of depression in the general population, the predictive power of MDD-PRS should be explored in other prospective stress samples. Second, although a PHQ-9 score of 10 and above is used widely as a cutoff for a depression diagnosis, it is important to note that the PHQ-9 is a screening tool and is susceptible to producing false-positives. Third, while the fivefold increase of depression within the first quarter of internship and steady depression rate throughout the year suggest that internship stress was the main driver of the dramatic increase in depression, normal age effect may have also contributed to the increase in depression during internship. Fourth, as the MDD-PRS incorporates input from across the genome, drawing definitive conclusions about underlying mechanism is not possible. Fifth, the polygenic scores described here were derived and tested in individuals of European ancestry only. Building up large enough samples to establish meaningful PRS for other ancestries is imperative to ensure that any benefits that follow from genomic medicine are not restricted to European ancestry populations. Lastly, the potential use of genetic risk disclosure should be balanced against potential harm. Specifically, as PRS improves, protections should be put in place to ensure that prediction cannot be used to discriminate against at-risk individuals. Reassuringly, our findings suggest that the MDD-PRS is better at identifying resilient individuals than individuals at most risk.

In summary, we find that MDD-PRS is a meaningful predictor of depression risk under stress. Future work should extend this work in multiple directions. First, the predictive power of MDD-PRS should be assessed in other prospective stress models, such as military stress and pregnancy. The similarities and differences in the genomic risk profiles across different types of stress would inform the extent to which future risk prediction work could be generalized across stressors or should be restricted to specific subtypes. Second, the finding that the percentage of the relationship between MDD-PRS and depression explained by established depression predictors decreased from 82% at baseline to 58% under internship stress suggests that there are important behavioural and psychological pathways from genomic risk to depression that are not fully explained. Further elucidating those pathways could enhance our understanding of the pathologic effects of stress. Similarly, the finding that individuals with extremely low MDD-PRS scores are particularly resilient to depression under stress suggests that prospective study of the behavioural and biological responses to stress among individuals with low MDD-PRS scores could identify strategies to prevent depression and inform the incorporation of precision medicine into psychiatry.

Methods

Participants. The Intern Health Study is a multi-institutional prospective cohort study that follows training physicians through the first year of residency training (internship). Interns entering residency programs across specialties in the academic years from 2007 to 2017 were sent an email 2–3 months before starting internship and invited to participate in the study. Subjects who consented to participate in the study were given a US$25 gift certificate after completing the baseline survey and another US$25 gift certificate after completing the follow-up survey[20,24].

The Institutional Review Board at the University of Michigan and the participating hospitals approved the study. All subjects provided informed consent after receiving complete description of the study.

Data collection. We measured depressive symptoms through the PHQ-9, a self-report component of the primary care evaluation of mental disorders inventory. For each of the nine depressive symptoms included in *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; ref. [27]), interns indicated whether, during the previous 2 wk, the symptom had bothered them 'not at all,' 'several days,' 'more than half the days' or 'nearly every day'. Each item yields a score of 0–3, so that the total score ranges from 0 to 27. The diagnostic validity of PHQ-9 has been shown to be comparable to clinician-administered assessments[28,29]. We used the PHQ-9 scores measured at baseline and during internship as the primary outcome measures for the study. PHQ depression, defined by a score of 10 or greater on the PHQ-9, which has moderate sensitivity (88%) and specificity (85%)[30] for a diagnosis of MDD was used as a secondary outcome in the study.

From 1 to 2 months before internship, subjects completed an online baseline survey, assessing PHQ-9 depressive symptoms, neuroticism (NEO-five factor inventory[31]), personal history of depression (self-reported yes/no) and early family environment (risky families questionnaire[32]), along with demographic information. We then contacted the participants via email at months 3, 6, 9 and 12 of their internship year and asked them to complete online surveys assessing PHQ-9 depressive symptoms and additional information about their internship experience. All surveys were conducted through a secure online website designed to maintain confidentiality, with subjects identified only by numeric IDs. No links between the identification number and the subjects' identities were maintained. Subjects who completed the baseline survey and at least one quarterly survey were included in the analysis, accounting for 86.08% of the total enroled subjects.

Genotyping and imputation. We collected DNA from subjects using DNA Genotek Oragene Mailable Tube (OGR-500)[33] through the mail. DNA ($n=9{,}611$) was extracted and genotyped on Illumina Infinium CoreExome−24 with v.1.0 or v.1.1 Chips, containing 571,054 and 588,628 SNPs, respectively.

We then implemented quality check of genotype data with PLINK v.1.9 (ref. [34]; www.cog-genomics.org/plink/1.9/). Samples with call rate < 99% ($n=179$) or with a sex mismatch between genotype data and reported data ($n=129$) were excluded. For 539 duplicated samples, the sample with higher call rate was selected. We only included SNPs on autosomal chromosomes, with call rate ≥ 98% (after sample removal) and minor allele frequency (MAF) ≥ 0.005. Genotype data from v.1.0 and v.1.1 Chips were then merged and subject duplication was again checked, and 11 more duplicates removed. A total 325,855 SNPs and 8,753 samples were considered for further analysis.

We performed LD-based pruning (window size 100 kilobases (kb), step size 25 variants, pairwise r^2 threshold 0.5) which yielded 202,235 SNPs for PCA of genotype data using all samples (PLINK v.1.9; ref. [34]). We defined European ancestry samples using the following steps. We plotted the first two PCs for self-reported European ancestry samples. On the basis of the plot, to reduce genetic heterogeneity, we included samples that fell within mean PC1 ± 3 s.d. and mean PC2 ± 6 s.d. We also included subjects who did not report ethnicity but whose PC1 and PC2 values were within the range defined above. From this set of European ancestry samples ($n=5{,}710$), we excluded SNPs with Hardy–Weinberg equilibrium $P<1\times10^{-6}$, leaving 325,249 genotyped SNPs. We defined East Asian ancestry ($n=816$) and South Asian ancestry ($n=595$) samples with the same procedure, except: for East Asian, the sample inclusion range was mean PC1 ± 3 s.d. and mean PC2 ± 4.5 s.d.; and for South Asian, it was mean PC1 ± 2.5 s.d. and mean PC2 ± 3.5 s.d. (Extended Data Fig. 3).

We performed genotype imputation on the Michigan Imputation Server using Minimac3 to phase samples and the 1000 Genomes Phase 3 data (v.5, phased by Eagle v.2.3) as reference panel.

MDD-PRS calculation. To calculate an MDD-PRS of MDD, we used the most recent MDD GWAS summary statistics from a meta-analysis of the PGC MDD phase 2, UK Biobank and 23andMe containing 246,363 cases and 561,190 controls (PGC phase 2 does not include Intern Health Study samples). We used PRSice v.2 (ref. [35]) to calculate MDD-PRS for our intern subjects. Following the recommendation from a recent review on PRS calculation[36], our primary analysis used the MDD-PRS calculated using the 213,863 variants genotyped in our sample (with MAF ≥ 0.1 and outside the major histocompatibility complex (MHC) region) that overlap with summary statistic data from the MDD GWAS. In a secondary analysis, we calculated MDD-PRS using imputed data (imputation quality > 0.9), first merging the imputed data with summary statistic data from the MDD GWAS, included only common SNPs with MAF ≥ 0.1 and then applying LD clumping (500 kb window and $r^2<0.1$). This process resulted in an MDD-PRS composed of 57,447 SNPs.

An additive model was used to calculate the PRS with the formula:

$$\mathrm{PRS} = \sum S \times G$$

where S is the MDD GWAS summary statistic effect size for the effect allele and $G=0,1,2$ is the number of effect alleles observed. The MDD-PRS was then mean-centred and scaled to 1 s.d.

Statistical analysis. Statistical analyses were conducted using R v.3.4.4 (The R Foundation) and SAS software v.9.4 for Windows (SAS Institute). We used 5,227 European ancestry samples with completed survey responses and genotype data for the main statistical analyses. We used an alpha level of 0.05 for all statistical tests. All tests were two-tailed. MDD-PRS, inverse-normalized PHQ-9 scores and covariates were assumed to be normal but normality was not formally tested. No statistical methods were used to predetermine sample size but our sample size is similar to those reported in previous publications[37–39].

Association of MDD-PRS with PHQ-9 depressive symptom score. Internship PHQ-9 depressive symptom score was calculated by averaging PHQ-9 score across all

available internship assessment. Raw PHQ-9 scores were left-skewed, so we applied inverse normalizing transformation on both baseline and internship PHQ-9 scores to produce near-normal distributions. Subjects who reported a PHQ-9 score greater than or equal to 10 at baseline were classified as meeting criteria for PHQ depression at baseline. Subjects who reported a PHQ-9 score ≥10 in at least one internship assessment were classified as meeting criteria for PHQ depression during internship. We used linear regression to test for association of MDD-PRS with inverse-normalized baseline PHQ-9 or internship PHQ-9, adjusting for age, sex and the top ten PCs of genotype data. We also used logistic regression to assess the association of MDD-PRS with baseline and internship PHQ depression.

To assess whether the effect size of MDD-PRS associating with PHQ-9 or PHQ depression was significantly different at baseline and during internship, we used a linear mixed model or a logistic mixed model to assess the interaction effect between MDD-PRS and internship stress, adjusting for main effect of MDD-PRS, internship stress, covariates including age, sex, top ten PCs of genotype data and the interaction terms for each covariate with MDD-PRS and internship stress. We included a random intercept term to account for correlation between repeated measurements within subjects.

We then conducted two sensitivity analyses: (1) to further assess the potential confounding effect of population stratification, we performed a sensitivity analysis by repeating the above linear regression models and linear mixed model with samples selected by a more conservative definition of European ancestry (PC1 < −0.002 and −0.005 < PC2 < 0.005); (2) to assess the robustness of our *P*-value threshold selection in PRS generation, we repeated the tests with MDD-PRS generated with *P*-value thresholds other than 1 (5×10^{-8}, 1×10^{-5}, 1×10^{-4}, 1×10^{-3}, 0.01, 0.05, 0.1 and 0.5).

Relationship between MDD-PRS and known risk factors of depression. In a separate analysis, we jointly included the three known risk factors (personal depression history, neuroticism score and early family environment)[20] as covariates in the models above. Continuous variables including neuroticism score and early family environment were scaled to have 1 s.d.

Mediation analysis was conducted to quantify the proportion of the MDD-PRS and PHQ-9 association mediated by known risk factors. R package 'mediation'[40] was used to fit our data in the following structural equation model:

$$Y = \alpha_1 + \beta_1 P + \varepsilon_1$$

$$M = \alpha_2 + \beta_2 P + \varepsilon_2$$

$$Y = \alpha_3 + \beta_3 P + \gamma M + \varepsilon_3$$

where Y is the outcome (inverse-normalized baseline or internship PHQ-9 or PHQ-9 change) and P is the predictor (MDD-PRS). Variable M is the mediator, either one of the known risk factors or the first PC of the three known risk factors. Variables α_i ($i=1,2,3$) and ε_i ($i=1,2,3$) represent the intercepts and errors, respectively. Coefficient β_1 represents the total effect of the MDD-PRS on PHQ-9, β_2 represents the effect of MDD-PRS on a known risk factor, β_3 represents the direct effect of MDD-PRS on PHQ-9 (the effect of MDD-PRS that is not accounted for by a known risk factor). Variable γ represents the effect of a known risk factor on PHQ-9. $\beta_2\gamma$ represents the indirect effect of MDD-PRS on PHQ-9 that is mediated by the known risk factor. The total effect of MDD-PRS on PHQ-9 is $\beta_3+\beta_2\gamma=\beta_1$. The mediated proportion $\beta_2\gamma/\beta_1$ is the proportion of the total MDR-PRS effect that is mediated by the known risk factor. To assess the significance of the mediated proportion, we performed 100,000 non-parametric bootstraps to estimate 95% CI and *P* value of the mediated proportion test for significant mediation, tested under the null hypothesis of the mediated proportion being zero.

Differences in PHQ-9 depressive symptoms or PHQ depression proportion by MDD-PRS percentile cutpoint. To test the ability of higher or lower MDD-PRS cutpoints to predict different levels of risk of PHQ depression (raw PHQ-9 score ≥ 10) or PHQ-9 scores, we serially dichotomized the sample based on an MDD-PRS percentile cutpoints of 2.5, 5, 10, 25, 75, 90, 95 and 97.5. At each MDD-PRS percentile cutpoint, we separate the subjects into two subgroups: (1) the tail subgroup with MDD-PRS percentile lower (for cutpoints < 50) or higher (for cutpoints > 50) than the cutpoint; (2) the remaining subgroup with all the other subjects. Then we compared the proportion of individuals with PHQ depression in the tail subgroup and the remaining subgroup using a Fisher's exact test. We also compared the PHQ-9 score in the tail subgroup and the remaining subgroup using a two sample *t*-test. For pairs of cutpoints that defined the same size of the sample at the low and high tails of the distribution (for example 2.5% and 97.5%), we used a permutation test to assess if the magnitude of the difference between the higher and lower MDD-PRS subgroups differed between the two cutpoints. Specifically, in each of 10,000 permutations, we permuted the MDD-PRS scores for all subjects, reran the Fisher's exact tests and *t*-tests for a given pair of cutpoints and calculated the ratio of the ORs ((1/low tail OR)/high tail OR) and the difference of two *t*-tests statistics ($-t$-statistic $_{\text{low tail MDD-PRS – remaining MDD-PRS}}$) − t-statistic $_{\text{high tail MDD-PRS – remaining MDD-PRS}}$). For each observed ratio of the odd ratios, we estimated the two-sided *P* value as the number of permutations with a ratio of ORs more extreme than either the observed ratio of the ORs or 1/(the observed ratio of ORs), divided by the number of permutations. For each difference in two *t*-tests, we calculated the two-sided *P* value as the number of permutations with an absolute (difference in two *t*-tests) > absolute (observed difference in the two *t*-tests), divided by the number of permutations.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The de-identified data from Intern Health Study are available through the PGC: https://www.med.unc.edu/pgc/shared-methods.
PGC phase 2–UK Biobank–23andMe MDD GWAS meta-analysis summary statistics: https://www.nature.com/articles/s41593-018-0326-7#data-availability.

Code availability

Custom code that supports the findings of this study is available from the corresponding author upon request.

Received: 6 March 2019; Accepted: 19 September 2019;
Published online: 28 October 2019

References

1. Friedrich, M. J. Depression is the leading cause of disability around the world. *J. Am. Med. Assoc.* **317**, 1517 (2017).
2. Rush, A. J. et al. Bupropion-SR, sertraline or venlafaxine-XR after failure of SSRIs for depression. *N. Engl. J. Med.* **354**, 1231–1242 (2006).
3. Rossen, L. M., Hedegaard, H., Khan, D. & Warner, M. County-level trends in suicide rates in the U.S., 2005–2015. *Am. J. Prev. Med.* **55**, 72–79 (2018).
4. Committee on Prevention of Mental Disorders & Institute of Medicine. *Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research* (National Academies Press, 1994).
5. O'Connell, M. E. et al. (eds) *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities* (National Academies Press, 2009).
6. Prendes-Alvarez, S. & Nemeroff, C. B. Personalized medicine: prediction of disease vulnerability in mood disorders. *Neurosci. Lett.* **669**, 10–13 (2018).
7. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *Am. J. Psychiatry* **157**, 1552–1562 (2000).
8. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
9. CONVERGE Consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588–591 (2015).
10. Hyde, C. L. et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).
11. Howard, D. M. et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat. Commun.* **9**, 1470 (2018).
12. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
13. 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).
14. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).
15. Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **17**, 392–406 (2016).
16. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
17. Kessler, R. C. The effects of stressful life events on depression. *Annu. Rev. Psychol.* **48**, 191–214 (1997).
18. Maciejewski, P. K. & Mazure, C. M. Stressful life events and depression. *Am. J. Psychiatry* **157**, 1344–1345 (2000).
19. Mata, D. A. et al. Prevalence of depression and depressive symptoms among resident physicians: a systematic review and meta-analysis. *J. Am. Med. Assoc.* **314**, 2373–2383 (2015).
20. Sen, S. et al. A prospective cohort study investigating factors associated with depression during medical internship. *Arch. Gen. Psychiatry* **67**, 557–565 (2010).
21. Martin, A. R. et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).

22. Navrady, L. B. et al. Genetic risk of major depressive disorder: the moderating and mediating effects of neuroticism and psychological resilience on clinical and self-reported depression. *Psychol. Med.* **48**, 1890–1899 (2018).
23. Middeldorp, C. M. & Wray, N. R. The value of polygenic analyses in psychiatry. *World Psychiatry* **17**, 26–28 (2018).
24. Guille, C. et al. Web-based cognitive behavioral therapy intervention for the prevention of suicidal ideation in medical interns: a randomized clinical trial. *JAMA Psychiatry* **72**, 1192–1198 (2015).
25. Southwick, S. M. & Charney, D. S. The science of resilience: implications for the prevention and treatment of depression. *Science* **338**, 79–82 (2012).
26. Harper, A. R., Nayee, S. & Topol, E. J. Protective alleles and modifier variants in human health and disease. *Nat. Rev. Genet.* **16**, 689–701 (2015).
27. *Diagnostic and Statistical Manual of Mental Disorders* 5th edn (American Psychiatric Association, 2003).
28. Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* **16**, 606–613 (2001).
29. Spitzer, R. L., Kroenke, K. & Williams, J. B. W. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *J. Am. Med. Soc.* **282**, 1737–1744 (1999).
30. Levis, B., Benedetti, A. & Thombs, B. D. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *Br. Med. J.* **365**, l1476 (2019).
31. Costa, P. T. Jr & McCrae, R. R. Stability and change in personality assessment: the revised NEO Personality Inventory in the year 2000. *J. Pers. Assess.* **68**, 86–94 (1997).
32. Taylor, S. E. et al. Early family environment, current adversity, the serotonin transporter promoter polymorphism, and depressive symptomatology. *Biol. Psychiatry* **60**, 671–676 (2006).
33. Rogers, N. L., Cole, S. A., Lan, H.-C., Crossa, A. & Demerath, E. W. New saliva DNA collection method compared to buccal cell collection techniques for epidemiological studies. *Am. J. Hum. Biol.* **19**, 319–326 (2007).
34. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
35. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2015).
36. Ware, E. B. et al. Heterogeneity in polygenic scores for common human traits. Preprint at *bioRxiv* https://doi.org/10.1101/106062 (2017).
37. Choi, K. W. et al. Prospective study of polygenic risk, protective factors, and incident depression following combat deployment in US Army soldiers. *Psychol. Med*. https://doi.org/10.1017/S0033291719000527 (2019).
38. Peyrot, W. J. et al. Effect of polygenic risk scores on depression in childhood trauma. *Br. J. Psychiatry* **205**, 113–119 (2014).
39. Musliner, K. L. et al. Polygenic risk, stressful life events and depressive symptoms in older adults: a polygenic score analysis. *Psychol. Med.* **45**, 1709–1720 (2015).
40. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. mediation: R package for causal mediation analysis. *J. Stat. Softw.* **59**, 1–38 (2014).

Acknowledgements

We thank the training physicians for taking part in this study. We also thank the participants and researchers of PGC and UK Biobank, and the research participants and employees of 23andMe, for making this work possible. This project was funded by the National Institute of Mental Health (grant no. R01MH101459 to S.S.). The funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

S.S. designed the study. S.S. and Y.F. developed the research question. Y.F. performed the data management and analysis. Y.F. and S.S. wrote the manuscript. L.S., P.S. and M.B. provided critical review, discussion and revision of the manuscript. All authors approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-019-0759-3.

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0759-3.

Correspondence and requests for materials should be addressed to S.S.

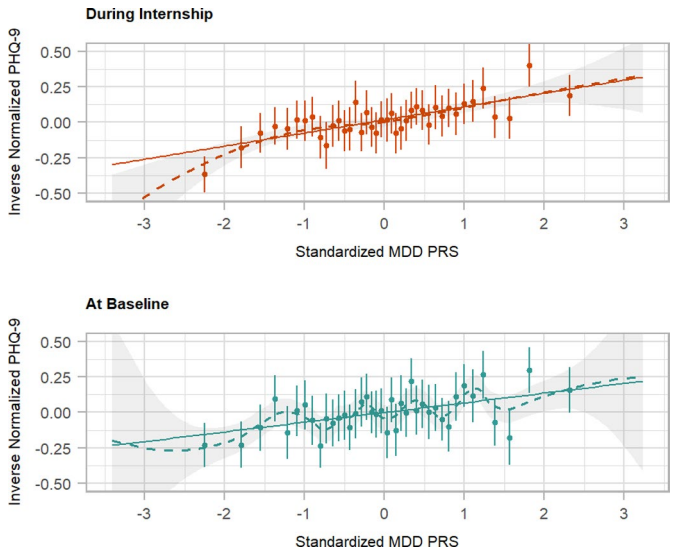
Peer review information Primary Handling Editor: Mary Elizabeth Sutherland.

Inclusion of known baseline risk factors* as covariates	Time Point of PHQ-9 score	Beta (SE)		p	
		MDD-PRS	MDD-PRS x internship stress Interaction**	MDD-PRS	MDD-PRS x internship stress Interaction**
Not included	Baseline	0.063 (0.014)		4.75×10^{-6}	
	Internship	0.092 (0.012)		3.18×10^{-15}	
	Baseline and internship	0.083 (0.034)	0.034 (0.013)	0.014	0.008
Included	Baseline	0.006 (0.012)		0.63	
	Internship	0.037 (0.010)		2.72×10^{-4}	
	Baseline and internship	0.026 (0.030)	0.033 (0.013)	0.38	0.010

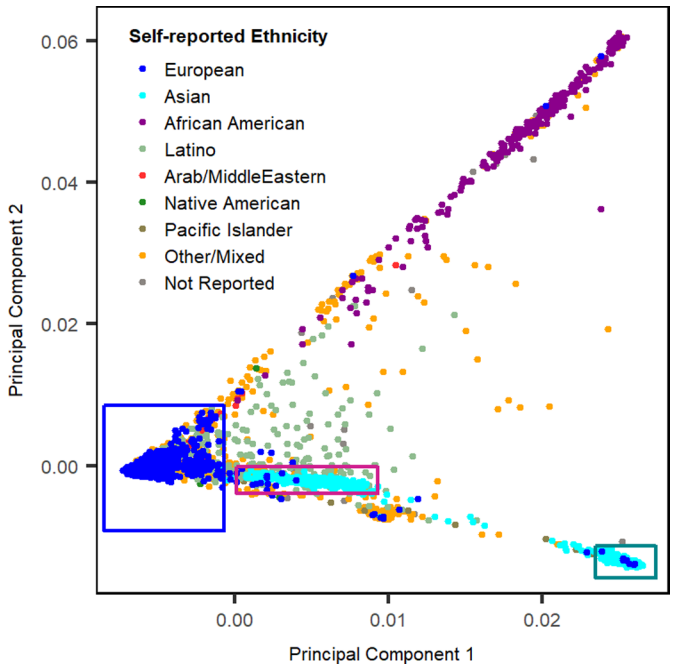
** Neuroticism, Personal History of Depression, Early Family Environment*

*** MDD-PRS x internship stress Interaction term is only available in the model including both baseline and internship data.*

Extended Data Fig. 1 | Associations of Imputed Data Derived MDD Polygenic Risk Score with PHQ-9 Depressive Symptom Scores (N = 5,227).



Extended Data Fig. 2 | Baseline and Internship PHQ-9 Depressive Symptom Scores by MDD-PRS Group. 5,227 Subjects from Intern Health Study were binned into 40 groups of 2.5% of subjects (n=131 per group) from low to high MDD-PRS (left to right). The 40 x-axis groups are defined by group-wise average standardized MDD-PRS. Average PHQ-9 score of each group at baseline (cyan dots) and during internship (orange dots) are plotted with 95% CI error bars. LOESS fitting line (dash line) shadowed by 95% CI and linear regression fitting line (solid line) were applied to both baseline and internship plots. Optimal span parameter for LOESS regression was selected by generalized cross-validation method.



Extended Data Fig. 3 | Population Structure Based on the Top Two Principal Component (PC) Analysis of the Intern Health Study. Blue, red and green boxes depicted the analysis inclusion range of European, South Asian and East Asian Groups.

nature research

Corresponding author(s): Srijan Sen

Last updated by author(s): Aug 5, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Qualtrics was used to collect survey data. GenomeStudio was to generate genotype data.
Data analysis	PRSice V2 was used to calculate MDD polygenic risk score. The statistic analysis was conducted by R (v3.4.4, packages: base packages, lme4, mediation, ggplot2) and SAS (v9.4, PROC GLIMMIX).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The de-identified data from Intern Health Study is available through the Psychiatric Genomics Consortium (PGC): https://www.med.unc.edu/pgc/shared-methods.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The prediction power and accuracy of polygenic risk score study is mostly dependent on the discovery sample. In this paper, the discovery sample is from PGC, UKB and 23andMe, and the sample size (246,363 cases and 561,190 controls) is the biggest for publicly major depression disorder GWAS so far. The sample size of 5,227 in the target sample (i.e. interns in our study) is sufficient achieve meaningful results in polygenic risk score study. Reference: Marees, A. T. et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int. J. Methods Psychiatr. Res. 27, e1608 (2018).
Data exclusions	Subjects was excluded when: no sufficient survey data, duplicated or low quality genotype data. The inclusion criteria of survey data followed previous publications from Intern Health Study. The quality check process of genotype data followed well-established and published protocols.
Replication	Replications were done: 1) in 5,116 subjects who fit a more conservative definition of European ancestry, and obtained fundamentally unchanged results; 2) using MDD-PRS generated with eight less inclusive p-value thresholds, the associations and interactions remained significant; 3) with imputed data derived MDD-PRS, and the results were slightly attenuated but still significant; 4) in East Asian and South Asian groups, and failed to produce significant results due to cross-ethnicity discrepancy in polygenic risk score prediction and smaller sample size.
Randomization	Randomization is not relevant to our study as there's no group allocation.
Blinding	Blinding is not relevant to our study as there's no group allocation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The population is first year residents (medial interns) in USA, with average age of 27.6 years old (SD = 2.7yrs), 50.3% females, and of European ancestry.
Recruitment	Interns entering residency programs across specialties in the academic years from 2007 to 2017 were sent an email 2-3 months prior to commencing internship and invited to participate in the study. Subjects consented to participate in the study were given a $25 gift certificate after completing the baseline survey and another $25 gift certificate after completing the follow-up survey. There is no bias present to impact the result.
Ethics oversight	The Institutional Review Board at the University of Michigan and the participating hospitals approved the study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.