Genome-wide association study implicates the role of TBXAS1 in the pathogenesis of depressive symptoms among the Korean population

Sungho Won (✉ won1@snu.ac.kr)

  Seoul National University   https://orcid.org/0000-0001-5751-5089

Kyungtaek Park

  https://orcid.org/0000-0003-0786-6969

Ah Ra Do

Yuree Chung

Min Ji Kim

  Seoul National University Hospital

Sang Jin Rhee

  Seoul National University College of Medicine   https://orcid.org/0000-0001-6221-5953

Dae Hyun Yoon

Seung Ho Choi

Sung Joon Cho

Han-Na Kim

Yong Min Ahn

  Seoul National University Hospital   https://orcid.org/0000-0002-4458-797X

Abstract

Although depression is an emerging disorder affecting many people worldwide, most genetic studies have been performed in European descent populations. Herein, a genome-wide association study (GWAS) was conducted in Korean population to elucidate the genomic loci associated with depressive symptoms. Two independent cohorts were used as discovery datasets, which consisted of 6 474 (1 484 cases and 4 990 controls) and 1 654 (557 cases and 1 097 controls) Korean participants, respectively. The participants were divided into case and control groups based on the Beck Depression Inventory (BDI). Meta-analysis using the two cohorts revealed that rs6945590 was significantly associated with the risk of depressive symptoms [P = 2.83 × 10^{-8}; odds ratio (OR) = 1.23; 95% confidence interval (CI): 1.15–1.33]. This association was validated in other independent cohorts which were another Korean cohort (258 cases and 1 757 controls) and the East Asian study of the Psychiatric Genomics Consortium (PGC) (15 771 cases and 178 777 controls). The predicted expression levels of *thromboxane A synthase 1* gene (*TBXAS1*), which encodes the enzyme thromboxane A synthase 1 and participates in the arachidonic acid (AA) cascade, was significantly decreased in the whole blood tissues of the participants with depressive symptoms. Furthermore, Mendelian randomization (MR) analysis showed a causal association between *TBXAS1* expression and the risk of depressive symptoms. In conclusion, as the number of risk alleles (A) of rs6945590 increased, *TBXAS1* expression decreased, which subsequently caused an increase in the risk of depressive symptoms.

Introduction

Depression is a common mental illness with symptoms, including sadness, lack of satisfaction, and sleep disturbance. Worldwide, approximately 280 million people suffer from depression in 2019, and the number only seems to increase from 1990 to 2019 [1]. The prevalence of depression in South Korea also has increased annually [2]. In particular, considering that depression is a major risk factor for suicide [3], the disease has been a big burden on South Korea. In 2019, the suicide rate recorded in South Korea was 24.6 per 100 000, which is the highest among the Organization for Economic Co-operation and Development countries [4].

Several loci associated with the risk of depression have been identified with efforts from many participants and researchers. For example, a meta-analysis with the highest number of participants to date showed that 102 loci are susceptible to the disease [5]. However, as the majority of genome-wide association study (GWAS) participants are of European descent [6], most depression studies have focused on the European populations [7]. Although a few independent variants have been reported in East Asian populations [8,9], more studies are needed, especially in non-European populations, to fully understand the etiology of the disease.

Another difficulty in shedding light on this disorder is that it is a polygenic trait; that is, its individual variants have only a small effect, which requires a large sample size to discover the susceptible loci. To overcome this problem, many GWASs have been performed comprising participants with self-reported depression [5,9–12], which is supported by the fact that self-declared depression and clinically diagnosed major depressive disorder (MDD) show a strong genetic correlation [11,13].

Here, we carried out GWAS to uncover variants associated with the risk of depressive symptoms in Korean population for the first time. Approximately 12 000 participants from three independent cohorts belonging to two different hospitals visited the institutes for a general health check-up. Moreover, their depression phenotypes were self-rated using Beck Depression Inventory (BDI) or Center for Epidemiologic Studies-Depression Scale (CES-D). Following GWAS, Mendelian randomization (MR) study was performed to identify the causal genomic loci.

Subjects And Methods

Health and Prevention Enhancement (H-PEACE) and Gene-Environment of Interaction and Phenotype (GENIE) cohorts

In the study of discovery stage, the H-PEACE and GENIE cohorts were used. They consisted of participants who visited the Seoul National University Hospital Healthcare System Gangnam Center for health check-ups from 2003 to 2017 [14]. Among them, 7 999 and 2 349 participants from the H-PEACE and GENIE cohorts, respectively, were genotyped on 833 000 variants using the Affymetrix Axiom Korean Chip (v1.0) [15]. The variant calling was performed using the K-Medoid clustering method [16].

Beck Depression Inventory

The degree of depression of the participants in the two cohorts was measured using original BDI. It is one of the most widely used self-scored measures of depression, comprising 21 items to evaluate how the participants had been feeling for the last two weeks [17]. Each item was rated on a four-point ordinal scale from 0 to 3, where a higher score represents a higher level of depression-related feelings. The total sum of the scores of 21 items was used as a criterion for depression.

In the cohorts, the BDI scores of a few participants had errors, because they had selected more than one point for an item or their responses for some items were missing. We regarded them as missing values and selected subjects who had missing values lower than 10.5, half of the total number of items. The average missing rates for each item in the H-PEACE and GENIE cohorts were 7.23% and 19.09%, respectively. We imputed missing values of the remaining participants by referring to (if any) their own scores measured during their check-ups at other times or scores of other participants who were close to them based on scores of other non-missing items using the missForest (v.1.4) R package [18]. In accordance with Beck et al. (1988) [19], we regarded participants

whose sum of the imputed scores was lower than 10 as controls and the others as cases. In the case of participants who had multiple health check-ups, the result was extracted when they had the highest BDI score (if they had the same scores, the latest result was extracted).

Genome-wide Association Study

After sample selection based on the BDI score, the genotype data were cleaned using the quality control steps suggested by Anderson et al (2010) [20]. Participants whose genotype missing rate was higher than 0.03, estimated heterozygosity rate differed from the sample mean of all participants by more than three standard deviations, and the estimated pairwise identity-by-descent was higher than 0.185 with a higher genotype missing rate than that of their counterparts were excluded. Variants whose genotype missing rate was significantly different between the case and control by Fisher's exact test ($P < 1 \times 10^{-5}$) and higher than 0.03, minor allele frequency (MAF) was lower than 0.05, and P-value of the Hardy–Weinberg equilibrium (HWE) test was lower than 1×10^{-3} were excluded (Supplementary Fig. 1).

Subsequently, the genotype data were imputed using the Michigan Imputation Server (v.1.1) [21]. Haplotype Reference Consortium r1.1 2016 and other/mixed population were chosen as reference panels, and phasing and imputation were performed using Eagle v2.4 and Minimac4, respectively. Among the imputed genotypes, variants whose MAF was higher than 0.05, genotype missing rate was lower than 0.03, HWE test was not significant at P-value 1×10^{-3}, and imputation quality measure "INFO" was higher than 0.8 were extracted (Supplementary Fig. 1).

Further, the GWAS was performed in each cohort after adjusting for the effects of sex, age, body mass index (BMI), and the top four principal component (PC) scores using plink (v1.90b3.44) [22]. For 67 and 18 (approximately 1%) participants in the H-PEACE and GENIE cohorts, respectively, BMIs were not observed and were imputed using missForest by referring to those of other participants who had similar sex and age [18]. The PC scores were calculated using pruned genotyped variants extracted by using *–indep-pairwise 50 5 0.2* option of plink. Meta-analysis was also performed with weights of inverse of the corresponding standard errors of each GWAS result using METAL (v2011-03-25) [23].

For genomic data analysis, in addition to plink (v1.90b3.44) and METAL (v2011-03-25), ONETOOL (v1.0) [24] and R (v3.6.3) were utilized.

Moreover, LocusZoom (http://locuszoom.org/) was used to generate a regional plot of ± 0.5 Mb flanking region of a significant variant [25]. Genome build and linkage disequilibrium (LD) population were selected as hg19 and 1000 Genomes Nov 2014 ASN, respectively. The other options were not manipulated.

Replication Cohorts

For validation, Kangbuk Samsung Cohort Study (KSCS) cohort was used. It is a prospective cohort study to evaluate the natural history, prognosis, and genetic and environmental determinants of a wide range of health traits and diseases among Korean adults. 2 011 participants were genotyped using the Illumina HumanCore BeadChips 12v. Depressive symptoms and severity were assessed using CES-D [26]. For identifying individuals at risk for clinical depressive symptoms, it uses 16 or above as cut-off score [27]. We divided the participants into two groups based on the cut-off score: depressed and non-depressed groups. The analysis steps used were the same as in H-PEACE and GENIE, except that 0.05 was used as the genotype missing rate cutoff in the participant QC step, and 0.02, 0.01, and 1×10^{-6} were used as the genotype missing rate, MAF, and HWE cutoffs, respectively, in the variant QC step. Furthermore, an older version of Minimac, Minimac3, was used in the imputation step.

Moreover, we referred an independent study of which meta-analysis of GWASs on depression in nine cohorts of East Asian ancestry [9]. The summary statistics, excluding the three cohorts, WHI, 23andME, and UKB, were downloaded from https://www.med.unc.edu/pgc/download-results/.

Expression Quantitative Trait Loci Analysis

LDexpress [28], a web-based tool implemented in LDlink [29] (https://analysistools.cancer.gov/LDlink/), was used to search for genes whose expression levels were associated with a significant variant from the GWAS results. In the East Asian population, genes in several tissues were identified, the expression levels of which were significantly associated ($P < 1 \times 10^{-5}$) with variantswhich were highly correlated ($R^2 > 0.8$) with a significant variant in ± 1 Mb window. The correlation between variants was estimated in the East Asian population using LDpair implemented in LDlink.

Transcriptome-wide Association Study

Gene expression levels in the whole blood tissues of the participants were predicted using PrediXcan [30]. The imputed genomic data of each cohort were used as input for the analysis. The gene-wise association between gene expression (a response variable) and self-rated depression (an explanatory variable) was estimated using a linear regression model with the same covariates used in the GWAS. Moreover, the false discovery rate (FDR) was estimated for each chromosome.

Mr Analysis

In this study, MR analysis was performed to identify the causal relationship between a gene and the risk of depressive symptoms. Generalized summary-data-based MR (GSMR) [31] implemented in Genome-wide complex trait analysis (GCTA) (v1.94.0b) [32] was used for the analysis with its default options. It required two summary statistics from regression models between variants and exposure or outcome and reference samples for LD estimation.

We estimated the summary statistic of the association between the expression levels of genes and variants in the imputed H-PEACE dataset using a linear regression model with the same covariates used in the GWAS. The GWAS result of the imputed GENIE dataset was used for the other summary statistic. East Asian samples in 1000 Genomes Phase 3 v5a were used as reference samples.

Association Study Between Variants Or A Gene And Each Depressive Symptom

GWASs of each item were performed considering the phenotype as 0 if the score was 0 or 1 otherwise and adjusting the effects of sex, age, BMI, and top four PC scores in each cohort. Meta-analyses for each item were performed using METAL. Associations between gene expression and the phenotype of each item were estimated with the same model used in the GWAS, except for replacing the number of minor alleles of a variant with the expression level of a gene in the two cohorts. Further, their Z statistics were summed by weighting their corresponding inverses of standard errors, and the P-values were calculated using the weighted Z statistics. Statistical analyses were performed using R and Rex (v3.6.1) [33].

Estimation Of Phenotypic And Genotype Correlations Between Depressive Symptoms

Based on the scores (0−3 scale), pairwise phenotypic correlations were estimated between the 21 BDI items in the H-PEACE and GENIE cohorts, respectively. Moreover, the items in combining the two cohorts were grouped into four clusters using *hcluster* function implemented in the amap R package (v0.8-18) with designating method and link options as Pearson and average, respectively. The phenotypes in each cluster were divided into two groups based on a cutoff in proportion to the number of items in the cluster. For example, the cutoff was set to 6.67, which was 10 times 14 divided by 21, in cluster 4 which included 14 items. GWASs were performed in each cluster of the two cohorts, respectively, with adjusting the effects of the same covariates used as above. Then, the results were combined by METAL.

In each cohort, genetic correlation between items were estimated by performing a bivariate GREML analysis implemented in GCTA with adjusting the effects of sex, age, BMI, and top four PC scores, respectively. Prevalence of each item was estimated as ratio of non-zero score of the item, and specified in the GREML analysis. Genetic relationship matrix was calculated by pruned variants with having larger than 0.1 MAF.

Ethics Statement

This study was approved by the institutional review boards of Seoul National University Hospital (H-1708-120-880) and Kangbuk Samsung Hospital (2019-08-006). All participants were provided written informed consent and the study was performed according to relevant guidelines and regulations.

Results

GWAS reveals that rs6945590 is associated with the risk of depressive symptoms

We performed a GWAS to elucidate the genetic loci susceptible to the risk of depressive symptoms in H-PEACE (4 233 391 variants in 1 484 cases and 4 990 controls) and GENIE (4 285 695 variants in 557 cases and 1 097 controls on) cohorts (Table 1). The proportion of females was higher in the cases than in the controls in both cohorts. Moreover, the age and BMI were similar between the cases and controls in both cohorts. Although several putative loci ($P < 1 \times 10^{-5}$) were identified, there were no significant variants at the conventional significance level ($P < 5 \times 10^{-8}$) in both cohorts (Supplementary Fig. 2 and Supplementary Tables 1 and 2). The most significant variants in the H-PEACE and GENIE cohorts were rs745327 (chr12:122456524) and rs16826069 (chr1:39797055), respectively.

Table 1
Participant characteristics and the number of variants in the H-PEACE and GENIE cohorts

	H-PEACE		GENIE	
	Case (N = 1,484)	Control (N = 4,990)	Case (N = 557)	Control (N = 1,097)
Age, mean (SD), yr	49.6 (10.9)	49.6 (9.7)	50.2 (10.1)	50.2 (9.3)
Women, N (%)	864 (58.2)	1,829 (36.7)	262 (47.0)	338 (30.8)
BMI, mean (SD), kg/m^{2*}	22.8 (3.2)	23.2 (2.9)	23.4 (3.2)	23.7 (2.9)
Variants after imputation, N	4,233,391		4,285,659	

Subsequently, a meta-analysis was performed using both cohorts. We analyzed a total of 4 141 113 variants in 8 128 participants (2 041 cases and 5 087 controls), of which one novel locus was identified (Fig. 1). At the locus, rs6945590 (chr7:139762282; 7q34) was significant ($P = 2.83 \times 10^{-8}$) at the conventional significance level (Fig. 1A). Moreover, the result was found to be reliable considering its quantile-quantile plot and a genomic inflation factor of 0.992 (Fig. 1B). P-values of the variants that were highly correlated with rs6945590 were lower than those of the variants that were not (Fig. 1C). In addition, rs6945590 is located in the intron region of the *poly(ADP-ribose) polymerase family member 12 (PARP12)* gene according to dbSNP (www.ncbi.nlm.nih.gov/snp). Moreover, several genes, such as the *thromboxane A synthase 1* gene *(TBXAS1)* are in close proximity to this gene (Fig. 1C).

Details regarding the GWAS results for rs6945590 are shown in Table 2. The risk and protective alleles of the variant were A and G, respectively. Moreover, the odds ratios (*ORs*) of the risk allele were 1.22 (95% confidence interval (CI): 1.12–1.33) and 1.26 (95% CI: 1.09–1.46) in H-PEACE and GENIE cohorts, respectively. The meta-analysis of both cohorts showed an OR of 1.23 (95% CI: 1.15–1.33). The risk allele frequencies (RAFs) of the cases were 0.608 and 0.584 in H-PEACE and GENIE cohorts, respectively. Moreover, the RAFs of the controls were 0.568 and 0.544 in H-PEACE and GENIE cohorts, respectively. Furthermore, these RAF values were found to be similar to those of the Korean population retrieved from Korean Reference Genome Database (KRGDB) (0.550) [34] and gnomAD (0.566) [35].

Table 2
Results of GWA studies and meta-analysis for rs6945590

Variants	Chr	BP	Risk / Protective alleles	H-PEACE			GENIE			Meta-analysis of H-PEACE and GENIE		KSCS		
				P-value	OR (95% CI)	RAF (Case / Control)	*P-value*	OR (95% CI)	RAF (Case / Control)	*P-value*	OR (95% CI)	*P-value*	OR (95% CI)	RAF (Case / Control)
rs6945590	7	139762282	A / G	3.79×10^{-6}	1.22 (1.12–1.33)	0.608 / 0.568	1.87×10^{-3}	1.26 (1.09–1.46)	0.584 / 0.544	2.83×10^{-8}	1.23 (1.15–1.33)	1.64×10^{-2}	1.28 (1.05–1.56)	0.409 / 0.360

Chr, chromosome; BP, base pair; OR, odds ratio; CI, confidence interval; RAF, risk allele frequency

We analyzed whether rs6945590 is associated with the risk of depressive symptoms or depression in other independent cohorts. In the KSCS cohort (4 792 372 variants in 258 cases and 1 752 controls), this association was validated at a nominal *P*-value 0.05 ($P = 1.64 \times 10^{-2}$; *OR* = 1.28; 95% *CI*: 1.05–1.56) (Table 2). When the meta-analysis of the H-PEACE, GENIE, and KSCS cohorts was performed, the association became more significant ($P = 1.59 \times 10^{-9}$; *OR* = 1.24; 95% *CI*: 1.16–1.33). Moreover, according to the other GWASs of depression in the East Asian population (13 163 200 variants in 15 771 cases and 178 777 controls) [9], rs6945590 was significantly associated with the risk of depression at a nominal *P*-value 0.05 ($P = 2.28 \times 10^{-2}$; *OR* = 1.04; 95% *CI*: 1.01–1.07).

In GWAS Catalog (www.ebi.ac.uk/gwas), traits that significantly associated with rs6945590 were not found, and in LDtrait (ldlink.nci.nih.gov/?tab = ldtrait), a web-based tool implemented in LDlink, association also was not reported between depression risk and rs6945590 or its highly correlated variants ($R^2 > 0.8$), however, some traits such as white blood cell count, monocyte count, and myocardial infarction had significant association with highly correlated variants with rs6945590 ($P < 5 \times 10^{-8}$; Supplementary Table 3).

TBXAS1 expression level is associated with rs6945590 and the risk of depressive symptoms

Genes for which rs6945590 was an eQTL were explored using LDexpress in the East Asian population. rs6945590 was found to be associated with *PARP12* expression in tibial nerve ($P = 2 \times 10^{-6}$) and testis ($P = 6 \times 10^{-6}$) tissues (Supplementary Table 4). Among pairs between genes and variants in the LD relationship with rs6945590 ($R^2 > 0.8$), *TBXAS1* and rs10242990 (chr7:139730500) pair in the whole blood tissue was the most significantly associated ($P = 9 \times 10^{-15}$; Supplementary Tables 5 and 6). rs10242990 and rs6945590 were highly correlated ($R^2 = 0.807$) and the risk allele (A) of rs6945590 was positively associated with the effect allele (G) of rs10242990. As the number of effect alleles of rs10242990 increased, *TBXAS1* expression levels decreased in the whole blood tissue ($\beta = -0.098$; Supplementary Table 5). *PARP12* also had eQTL variants in the LD region of rs6945590, but its expression levels were not significantly associated with the variants in the whole blood or brain tissues at a nominal *P*-value 1×10^{-5} (Supplementary Table 6).

We predicted the expression levels of genes in the whole blood tissue of the participants from the genotypes of the H-PEACE and GENIE cohorts using PrediXcan to identify the genes associated with the risk of depressive symptoms (Fig. 2). In the H-PEACE cohort, the predicted *TBXAS1* expression level was negatively significantly associated with the risk of depressive symptoms at an FDR of 0.05 ($\beta = -0.023$; $P = 3.54 \times 10^{-5}$; FDR = 1.26×10^{-2}) (Fig. 2A), and the association was validated in the GENIE cohort at a nominal *P*-value 0.05 ($\beta = -0.023$; $P = 2.10 \times 10^{-2}$) (Fig. 2B). *TBXAS1* expression level was also significantly associated with rs6945590 in both cohorts. As the number of risk alleles (A) of rs6945590 increased, *TBXAS1* expression decreased (Fig. 2C and 2D).

Moreover, *lanosterol synthase* (*LSS*) and *YbeY metalloendoribonuclease* (*YBEY*) may be other putative genes associated with the risk of depressive symptoms (Supplementary Fig. 3). *LSS* was predicted to be significantly more expressed in the cases than in the controls in the H-PEACE cohort ($\beta = 0.046$; $P = 1.43 \times 10^{-4}$; FDR = 6.30×10^{-3}) and the same pattern was observed in the GENIE cohort ($\beta = 0.033$; one tail $P = 4.28 \times 10^{-2}$) (Supplementary Fig. 3A and 3B; and Supplementary Table 7). The predicted *YBEY* expression levels significantly decreased in the cases compared to the controls in the H-PEACE cohort at an FDR of 0.1 ($\beta = -0.031$; $P = 3.40 \times 10^{-3}$; FDR = 7.48×10^{-2}), and it was validated in the GENIE cohort ($\beta = -0.038$; $P = 4.92 \times 10^{-2}$) (Supplementary Fig. 3C and 3D; and Supplementary Table 7). Although the genes ENSG00000235878, *4-hydroxyphenylpyruvate dioxygenase*, *cysteine and tyrosine rich 1*, and *spermatogenesis and centriole associated 1* like showed a significant association with the risk of depressive symptoms in the H-PEACE cohort, they did not in the GENIE cohort (Supplementary Table 7).

MR study shows that TBXAS1 may be a causal gene of depressive symptoms

We performed a two-sample MR study using the GSMR to test whether genes and the risk of depressive symptoms had a causal relationship. The two genome-wide association results were from the analyses: (1) between variants and *TBXAS1* expression in the H-PEACE cohort and (2) between variants and

the risk of depressive symptoms in the GENIE cohort. With 13 independent instrumental variables (IVs; $R^2 < 0.05$; $P < 5.00 \times 10^{-8}$), the study revealed that a decrease in *TBXAS1* expression led to an increase in the risk of depressive symptoms ($\beta = -0.785$; $P = 3.82 \times 10^{-3}$) (Fig. 3). When the *P*-value cutoff for IV selection was increased to 1.00×10^{-5}, the effect was still significant, with 23 IVs ($\beta = -0.783$; $P = 3.42 \times 10^{-3}$) (Supplementary Fig. 4A). The reverse causation could not be tested because there were not enough variants available as IVs even when the cutoff was set to 1.00×10^{-5}.

The causal relationships between *LSS* or *YBEY* and the risk of depressive symptoms were also tested. Although there were not enough IVs when the cutoff was set to the conventional threshold, 5.00×10^{-8}, for both directions, *LSS* may be a putative causal gene for depressive symptoms considering the result when the cutoff was 1.00×10^{-5} with 17 IVs ($\beta = 0.280$; $P = 3.51 \times 10^{-2}$) (Supplementary Fig. 4B). If *LSS* expression increases, the risk of depressive symptoms may increase. *YBEY* may also have a causal relationship with the risk of depressive symptoms. A decrease in its expression appeared to increase the risk of depressive symptoms, but its nominal *P*-value was slightly higher than 0.05 ($\beta = -0.265$; $P = 6.99 \times 10^{-2}$). It was estimated with 10 IVs that were selected at the conventional cutoff (Supplementary Fig. 4C). There were not enough variants available as IVs to test the reverse causation even when the cutoff was 1.00×10^{-5} in both genes.

Some Symptoms Of Depression Are More Highly Associated With Rs6945590 Than Others

The associations between rs6945590 or *TBXAS1* expression and each item were investigated. As the number of risk alleles of rs6945590 increased, the ratios of non-zero value of 12 out of the 21 items (items 1, 2, 4, 5, 7, 8, 12, 13, 14, 15, 16, and 20) significantly increased at an FDR of 0.05 in (Fig. 4). Among them, items 4 (lack of satisfaction) and 16 (sleep disturbance) had the smallest *P*-values. *TBXAS1* expression was significantly associated with the items 4 and 16 at an FDR of 0.05 (Fig. 4).

Scores for each item were significantly correlated with each other except for item 19 (weight loss). Moreover, the scores for items 9 (suicide) and 10 (cry) showed the highest correlation in both the H-PEACE and GENIE cohorts (Supplementary Fig. 5A). We clustered the items based on their pairwise score correlations (Supplementary Fig. 5B). There were four clusters: cluster 1 included item 19; cluster 2 included item 6; cluster 3 included items 1, 2, 3, 9, and 10; cluster 4 included the other remaining 14 items. 10 of 12 items which had significant association with rs6945590 were included in the cluster 4.

GWASs of each cluster showed that rs6945590 was significantly associated with the risk of depressive symptoms in only cluster 4 ($P = 1.95 \times 10^{-8}$; $OR = 1.22$; 95% *CI*: 1.14−1.31; Supplementary Table 8). Moreover, there was no other significant variant in the four clusters (Supplementary Fig. 6).

Pairwise genetic correlations between items in each cohort were not significantly different from 0 at an FDR of 0.05 except items 5 (guilt) and 7 (self-hate) in GENIE cohort which was not significant in H-PEACE cohort (Supplementary Fig. 5C). Furthermore, GWASs were performed for each item using the same procedure as above. However, there was no significant variant at the 5.00×10^{-8} significance level (Supplementary Fig. 7).

Discussion

In this study, rs6945590 was identified as a novel susceptible variant of depressive symptoms in the Korean population (Fig. 1 and Table 2). The variant was shown to be significantly associated with the risk of depressive symptoms or depression in three independent cohorts consisting of Korean and one cohort consisting of East Asian. As the number of risk alleles of rs6945590 increased, the expression level of *TBXAS1* decreased (Fig. 2), subsequently increasing the risk of depressive symptoms (Fig. 3). Among symptoms of depression, lack of satisfaction and sleep disturbance were highly associated with rs6945590 and *TBXAS1* (Fig. 4).

Other than rs6945590, no variants were significantly associated with risk of depressive symptoms in our study including variants which were shown to be in European study [5]. Although 59 out of 102 significant variants in the European study existed in our dataset, only 2 variants (rs113188507 and rs9592461) had the same direction of OR with significant association with risk of depressive symptoms at nominal *P*-value 0.05 in our study (Supplementary Table 9). It may implicate importance of our study to fill missing heritability of depression.

Patients with depression have not only heterogeneous symptoms, but also different drug responses. Only a third of the patients are treated by the currently developed drugs that largely target the monoamine pathway; the other half are partly cured by the drugs, and the rest do not respond to them [36,37]. One promising remedy for this unmet need is to utilize the relationship between inflammation and neuropsychiatric diseases. In recent years, emerging evidence has shown that immune mechanisms can contribute to depression [38,39]. For example, higher interleukin-6 levels in blood precede depression [40,41] and anti-inflammatory treatments such as non-steroidal anti-inflammatory drugs (NSAIDs) show antidepressant effects [42].

The arachidonic acid (AA) cascade is an immune response pathway that plays a key role in many inflammatory diseases such as asthma [43]. AA is converted to prostaglandin H2 (PGH2), catalyzed by cyclooxygenase (COX), and subsequently to prostaglandins or thromboxane A2 (TXA2). In the latter step, thromboxane A synthase 1, encoded by *TBXAS1*, catalyzes the conversion of PGH2 to TXA2, which is known as an inducer of platelet aggregation.

Considering that TXA2 activates the immune response, which may be positively associated with the risk of depression, it appears that the risk of depressive symptoms may decrease with the decrease in *TBXAS1* expression. However, in the AA to TXA2 cascade, TXA2 itself may have a limited effect on depressive symptoms. However, the amount of AA appears to have an important effect on depression. Several studies have shown that the ratio of AA to long-chain omega-3 fatty acids is elevated in depression [44,45]. Moreover, continued use of non-aspirin NSAIDs or high-dose aspirin (COX inhibitors that block the conversion of AA to PGH2) increases the incidence rate of depression [46]. Moreover, aspirin aggravates depressive symptoms [47,48].

In addition to COX, AA is metabolized by lipoxygenases (LOXs). LOXs catalyze the conversion of AA to leukotriene, which has broad proinflammatory effects [49]. For example, knockdown of the leukotriene receptor prevents depressive behavior in mice [50]. Moreover, leukotriene receptor antagonists were reported to be associated with sleep disorder [51, 52]. Leukotriene is mainly synthesized by leukocytes and monoblastic lineages such as monocytes [53]. In addition, according to the Human Protein Atlas (https://www.proteinatlas.org/), *TBXAS1* is more highly expressed in monocytes than in other cell types in peripheral blood mononuclear cells.

LDtrait showed that rs6945590 was highly correlated with the variants that were significantly associated with the immune cell counts in the East Asian population (Supplementary Table 3). rs2269996 (chr7:139724358) was significantly associated with the white blood cell and monocyte counts. rs7781964 (chr7:139749346) was significantly associated with the lymphocyte count. Moreover, studies have shown that white blood cell count and monocyte count are positively associated with depression risk [54, 55].

Therefore, we suggest that a decrease in *TBXAS1* expression causes the accumulation of AA, which subsequently activates the immune response through leukotrienes by leukocytes or monocytes, and it leads to increase the risk of depressive symptoms.

However, our findings must be scrutinized by further research because of the following limitations: First, genomic analyses were based on imputed data. rs6945590 was an imputed variant and not a genotyped variant; gene expression levels were predicted, and missing values of each item were interpolated. Although our findings should be demonstrated with complete data, such as whole genome sequencing and mRNA sequencing, we believe that they are not false positives because they were validated in independent datasets. Significant associations between rs6945590 or *TBXAS1* expression and the risk of depressive symptoms or depression were found in at least two independent cohorts. Second, the study did not have a sufficient sample size; therefore, we could not find the genomic loci associated with each heterogeneous symptom of depression and other variants that have relatively small effects on depressive symptoms in Koreans. Moreover, we may not have enough power to estimate genetic correlation between the symptoms. Further studies with larger sample sizes are needed to fully understand the genetic mechanism of depression and to determine whether *TBXAS1* could be used as a blood marker or target for treating depression.

In conclusion, a new causal genomic locus of depressive symptoms was identified in the Korean population. rs6945590 and *TBXAS1* were significantly associated with the risk of depressive symptoms. *TBXAS1* might play a role in the pathogenesis of depressive symptoms by affecting immune response. These findings might help to find novel markers or therapeutics for the diagnosis or treatment of depressive symptoms.

Declarations

Conflict of Interest

The authors declare no conflict of interest.

References

1. Global Health Data Exchange (GHDx). http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/d780dffbe8a381b25e1416884959e88b, Accessed 4 July 2022.
2. Kim GE, Jo M-W, Shin Y-W. Increased prevalence of depression in South Korea from 2002 to 2013. Scientific reports 2020; **10**(1): 1−9.
3. Chesney E, Goodwin GM, Fazel S. Risks of all-cause and suicide mortality in mental disorders: a meta-review. World psychiatry 2014; **13**(2): 153−160.
4. OECD. *Health at a Glance 2021*, 2021.
5. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirali M *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nature neuroscience 2019; **22**(3): 343−352.
6. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature 2016; **538**(7624): 161−164.
7. Peterson RE, Kuchenbaecker K, Walters RK, Chen C-Y, Popejoy AB, Periyasamy S *et al.* Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. Cell 2019; **179**(3): 589−603.
8. Cai N, Bigdeli TB, Kretzschmar W, Li Y, Liang J, Song L *et al.* Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature 2015; **523**(7562): 588−591.
9. Giannakopoulou O, Lin K, Meng X, Su M-H, Kuo P-H, Peterson RE *et al.* The genetic architecture of depression in individuals of East Asian ancestry: a genome-wide association study. JAMA psychiatry 2021; **78**(11): 1258−1269.
10. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nature genetics 2016; **48**(9): 1031−1036.
11. Howard DM, Adams MJ, Shirali M, Clarke T-K, Marioni RE, Davies G *et al.* Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. Nature communications 2018; **9**(1): 1−10.
12. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nature genetics 2018; **50**(5): 668−681.

13. Thorp JG, Marees AT, Ong J-S, An J, MacGregor S, Derks EM. Genetic heterogeneity in self-reported depressive symptoms identified through genetic analyses of the PHQ-9. Psychological Medicine 2020; **50**(14): 2385–2396.

14. Lee C, Choe EK, Choi JM, Hwang Y, Lee Y, Park B *et al.* Health and Prevention Enhancement (H-PEACE): a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea. BMJ open 2018; **8**(4): e019327.

15. Moon S, Kim YJ, Han S, Hwang MY, Shin DM, Park MY *et al.* The Korea Biobank Array: design and identification of coding variants associated with blood biochemical traits. Scientific reports 2019; **9**(1): 1–11.

16. Seo S, Park K, Lee JJ, Choi KY, Lee KH, Won S. SNP genotype calling and quality control for multi-batch-based studies. Genes & genomics 2019; **41**(8): 927–939.

17. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Archives of general psychiatry 1961; **4**(6): 561–571.

18. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics 2012; **28**(1): 112–118.

19. Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. Clinical psychology review 1988; **8**(1): 77–100.

20. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nature protocols 2010; **5**(9): 1564–1573.

21. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A *et al.* Next-generation genotype imputation service and methods. Nature genetics 2016; **48**(10): 1284–1287.

22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics 2007; **81**(3): 559–575.

23. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010; **26**(17): 2190–2191.

24. Song YE, Lee S, Park K, Elston RC, Yang H-J, Won S. ONETOOL for the analysis of family-based big data. Bioinformatics 2018; **34**(16): 2851–2853.

25. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP *et al.* LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010; **26**(18): 2336–2337.

26. Cho MJ, Kim KH. Use of the Center for Epidemiologic Studies Depression (CES-D) Scale in Korea. J Nerv Ment Dis 1998; **186**(5): 304–310.

27. Weissman MM, Sholomskas D, Pottenger M, Prusoff BA, Locke BZ. Assessing depressive symptoms in five psychiatric populations: a validation study. Am J Epidemiol 1977; **106**(3): 203–214.

28. Lin S-H, Thakur R, Machiela MJ. LDexpress: an online tool for integrating population-specific linkage disequilibrium patterns with tissue-specific expression data. BMC bioinformatics 2021; **22**(1): 1–6.

29. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 2015; **31**(21): 3555–3557.

30. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ *et al.* A gene-based association method for mapping traits using reference transcriptome data. Nature genetics 2015; **47**(9): 1091–1098.

31. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. Nature communications 2018; **9**(1): 1–12.

32. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. The American Journal of Human Genetics 2011; **88**(1): 76–82.

33. RexSoft (2018). Rex: Excel-based statistical analysis software. URL http://rexsoft.org/.

34. Jung KS, Hong K-W, Jo HY, Choi J, Ban H-J, Cho SB *et al.* KRGDB: the large-scale variant database of 1722 Koreans based on whole genome sequencing. *Database* 2020; **2020**.

35. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020; **581**(7809): 434–443.

36. Warden D, Rush AJ, Trivedi MH, Fava M, Wisniewski SR. The STAR* D Project results: a comprehensive review of findings. Current psychiatry reports 2007; **9**(6): 449–459.

37. Miller AH, Maletic V, Raison CL. Inflammation and its discontents: the role of cytokines in the pathophysiology of major depression. Biological psychiatry 2009; **65**(9): 732–741.

38. Arteaga-Henríquez G, Simon MS, Burger B, Weidinger E, Wijkhuijs A, Arolt V *et al.* Low-grade inflammation as a predictor of antidepressant and anti-inflammatory therapy response in MDD patients: a systematic review of the literature in combination with an analysis of experimental data collected in the EU-MOODINFLAME consortium. Frontiers in psychiatry 2019; **10**: 458.

39. Branchi I, Poggini S, Capuron L, Benedetti F, Poletti S, Tamouza R *et al.* Brain-immune crosstalk in the treatment of major depressive disorder. European Neuropsychopharmacology 2021; **45**: 89–107.

40. Khandaker GM, Pearson RM, Zammit S, Lewis G, Jones PB. Association of serum interleukin 6 and C-reactive protein in childhood with depression and psychosis in young adult life: a population-based longitudinal study. JAMA psychiatry 2014; **71**(10): 1121–1128.

41. Lamers F, Milaneschi Y, Smit JH, Schoevers RA, Wittenberg G, Penninx BW. Longitudinal association between depression and inflammatory markers: results from the Netherlands study of depression and anxiety. Biological psychiatry 2019; **85**(10): 829–837.

42. Drevets WC, Wittenberg GM, Bullmore ET, Manji HK. Immune targets for therapeutic development in depression: towards precision medicine. Nature Reviews Drug Discovery 2022: 1–21.

43. Wang B, Wu L, Chen J, Dong L, Chen C, Wen Z *et al.* Metabolism pathways of arachidonic acids: Mechanisms and potential therapeutic targets. Signal transduction and targeted therapy 2021; **6**(1): 1–30.

44. Lotrich FE, Sears B, McNamara RK. Elevated ratio of arachidonic acid to long-chain omega-3 fatty acids predicts depression development following interferon-alpha treatment: relationship with interleukin-6. Brain, behavior, and immunity 2013; **31**: 48–53.

45. Adams PB, Lawson S, Sanigorski A, Sinclair AJ. Arachidonic acid to eicosapentaenoic acid ratio in blood correlates positively with clinical symptoms of depression. Lipids 1996; **31**(1Part2): S157-S161.

46. Kessing LV, Rytgaard HC, Gerds T, Berk M, Ekstrøm C, Andersen P. New drug candidates for depression–a nationwide population-based study. Acta Psychiatrica Scandinavica 2019; **139**(1): 68–77.

47. Berk M, Mohebbi M, Dean OM, Cotton SM, Chanen AM, Dodd S *et al.* Youth Depression Alleviation with Anti-inflammatory Agents (YoDA-A): a randomised clinical trial of rosuvastatin and aspirin. BMC medicine 2020; **18**(1): 1–12.

48. Berk M, Agustini B, Woods RL, Nelson MR, Shah RC, Reid CM *et al.* Effects of aspirin on the long-term management of depression in older people: a double-blind randomised placebo-controlled trial. Molecular psychiatry 2021; **26**(9): 5161–5170.

49. Peters-Golden M, Canetti C, Mancuso P, Coffey MJ. Leukotrienes: underappreciated mediators of innate immune responses. The Journal of Immunology 2005; **174**(2): 589–594.

50. Yu X-B, Dong R-R, Wang H, Lin J-R, An Y-Q, Du Y *et al.* Knockdown of hippocampal cysteinyl leukotriene receptor 1 prevents depressive behavior and neuroinflammation induced by chronic mild stress in mice. Psychopharmacology 2016; **233**(9): 1739–1749.

51. Cereza G, Doladé NG, Laporte J-R. Nightmares induced by montelukast in children and adults. European Respiratory Journal 2012; **40**(6): 1574–1575.

52. Hara H, Sugahara K, Hashimoto M, Mikuriya T, Tahara S, Yamashita H. Effectiveness of the leukotriene receptor antagonist pranlukast hydrate for the treatment of sleep disorder in patients with perennial allergic rhinitis. Acta Oto-Laryngologica 2014; **134**(3): 307–313.

53. Haeggström JZ. Leukotriene biosynthetic enzymes as therapeutic targets. The Journal of clinical investigation 2018; **128**(7): 2680–2690.

54. Seidel A, Arolt V, Hunstiger M, Rink L, Behnisch A, Kirchner H. Major depressive disorder is associated with elevated monocyte counts. Acta Psychiatrica Scandinavica 1996; **94**(3): 198–204.

55. Sealock JM, Lee YH, Moscati A, Venkatesh S, Voloudakis G, Straub P *et al.* Use of the PsycheMERGE Network to Investigate the Association Between Depression Polygenic Scores and White Blood Cell Count. JAMA psychiatry 2021; **78**(12): 1365–1374.
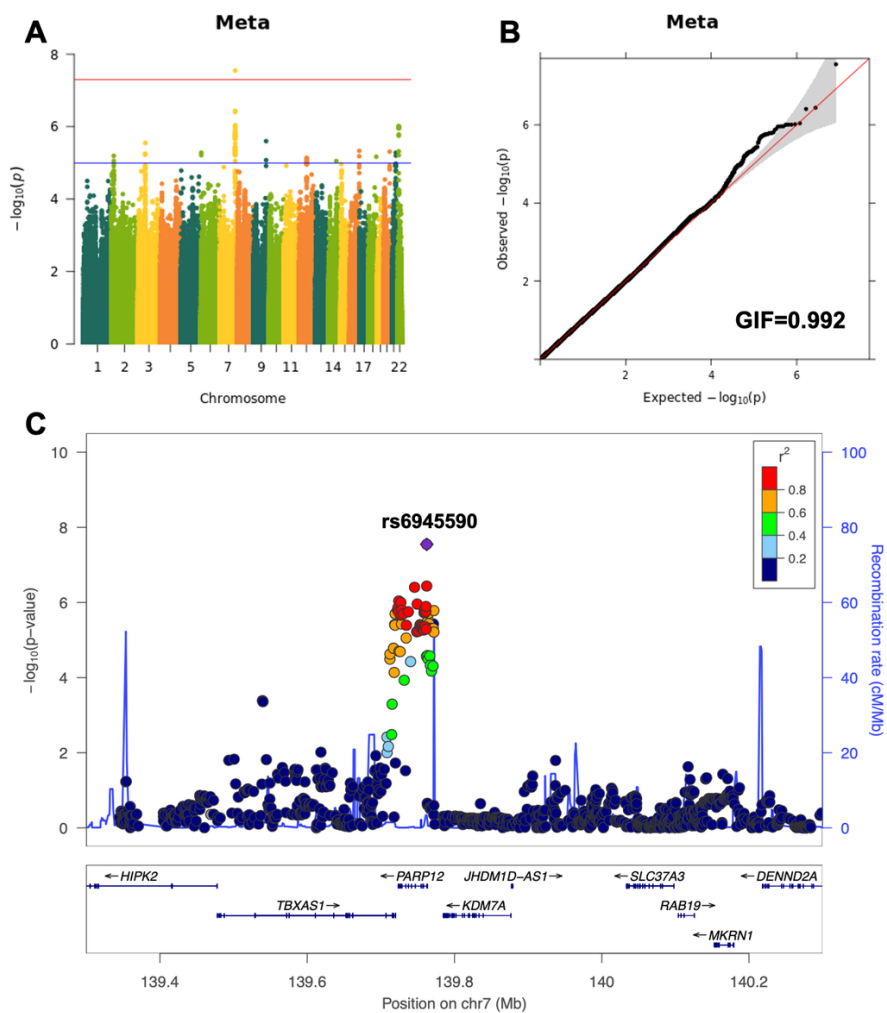
Figures

Figure 1

Meta-analysis results of the risk of depressive symptoms. (A) Manhattan plot. Red and blue lines represent significance level 5.00×10^{-8} and 1.00×10^{-5}, respectively. (B) Quantile–quantile plot. Grey region indicates point-wise 95% confidence interval (CI) of P-values. (C) Regional plot. Purple square represents rs6945590 and circles represent other variants. Colors represent degree of correlation (r^2) with rs6945590, and genes near the variant are shown below. p, P-value; GIF, genomic inflation factor; chr, chromosome.
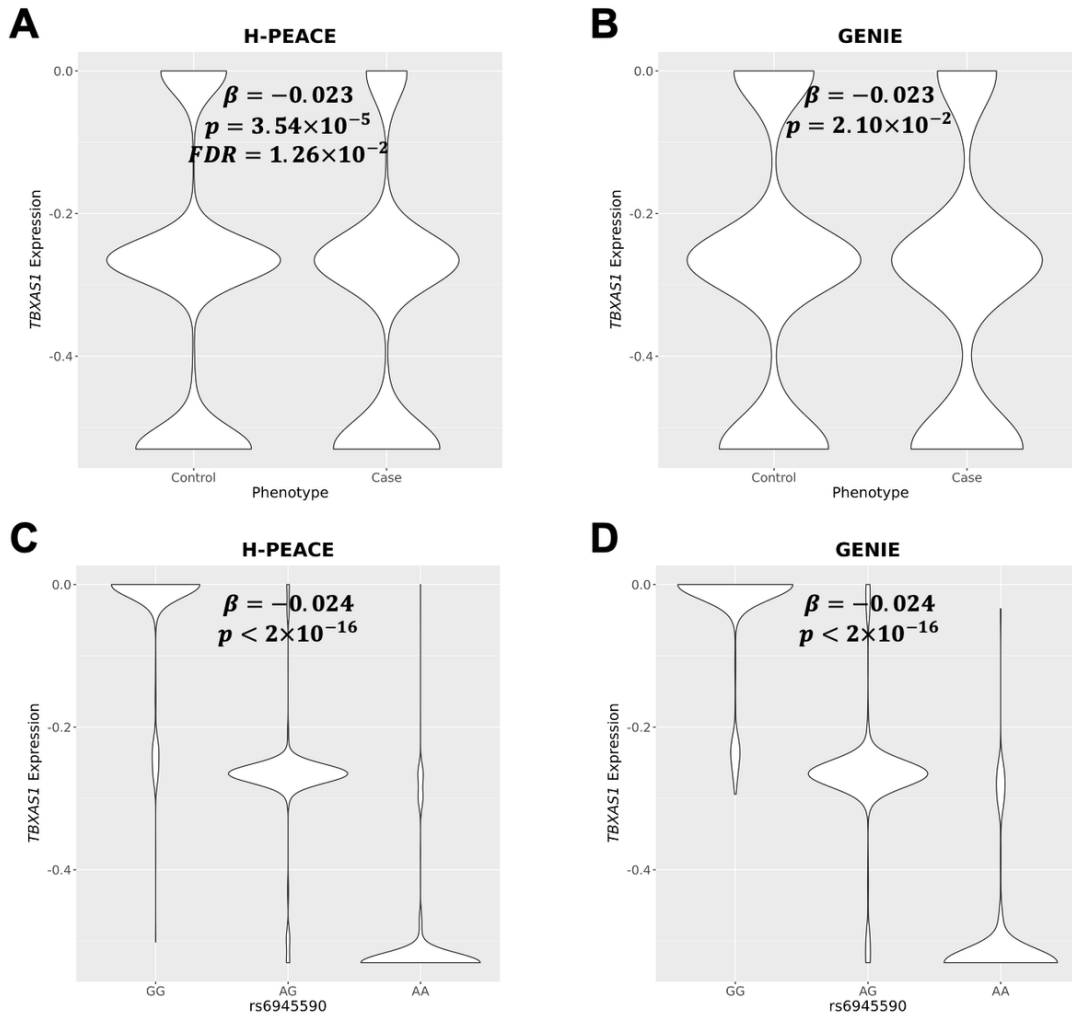
Figure 2

Association between *TBXAS1* expression levels and risk of depressive symptoms or rs6945590 in whole blood. Violin plots of predicted *TBXAS1* expression levels according to the depression phenotypes of the (A) health and prevention enhancement (H-PEACE) and (B) gene-environment of interaction and phenotype (GENIE) cohorts and alleles of rs6945590 of (C) H-PEACE and (D) GENIE cohorts. *β*, estimate of the association between *TBXAS1* expression and rs6945590 or the phenotypes; p, *P*-value; FDR, false discovery rate.
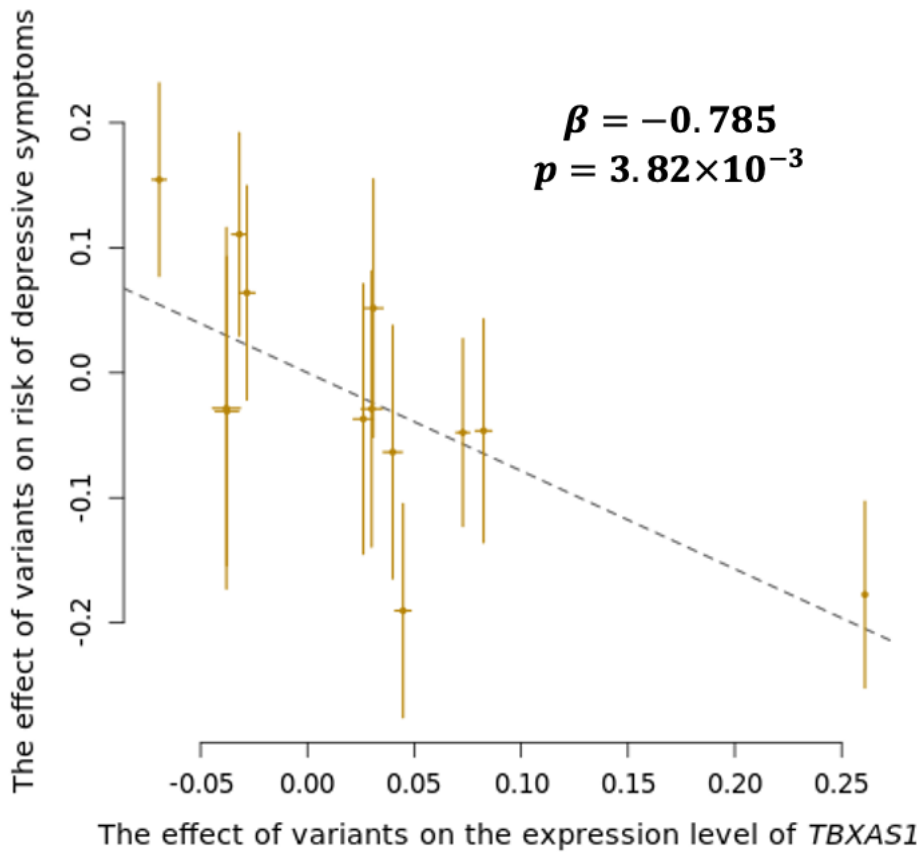
$$\beta = -0.785$$
$$p = 3.82 \times 10^{-3}$$

Figure 3

Causal effect of *TBXAS1* expression on the risk of depressive symptoms. Effects of 13 instrumental variables (IVs) are shown in yellow dots with their ±1 standard error (yellow lines). Slope of dashed lines represents the causal effect of *TBXAS1* expression on the risk of depressive symptoms. *β*, estimate of the causal effect; p, *P*-value.
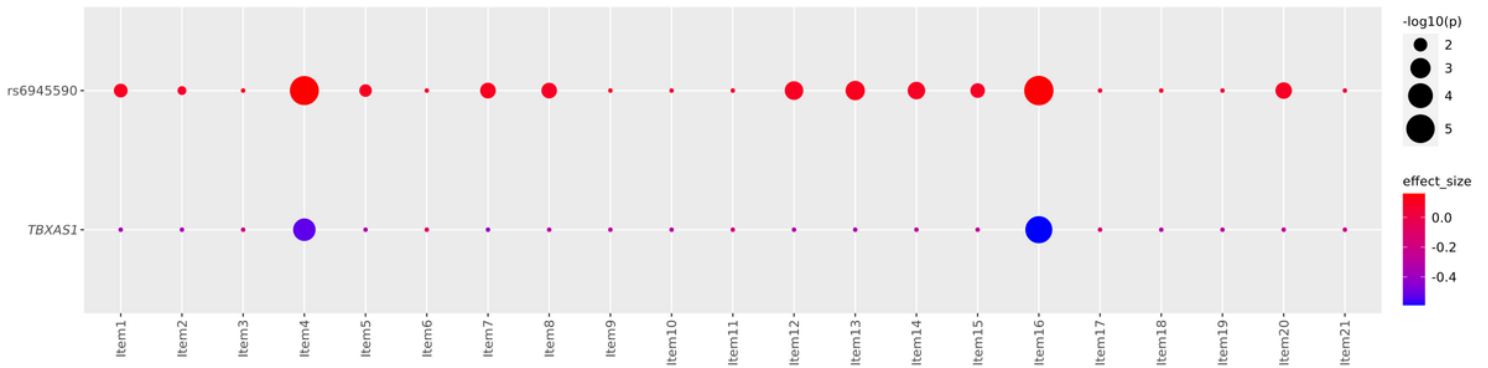

Figure 4

Associations between each item and rs6945590 or *TBXAS1*. Effect sizes are represented by circles with colors ranging from red (positive) to blue (negative). Size of the circle represents *P*-values expressed as -log10 scale. Circles with the smallest size indicate that the association is not significant. 12 and 2 out of the 21 items were significantly associated with rs6945590 and *TBXAS1*, respectively. p, *P*-value.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.docx
- SupplementaryTables.docx