DNA methylation signatures of depressive symptoms in middle-aged and elderly persons: meta-analysis of multi-ethnic epigenome-wide studies

Olivera Story Jovanova*, Ivana Nedeljković*, Derek Spieler*, Rosie M. Walker*, Chunyu Liu*, Michelle Luciano*, Jan Bressler, Jennifer Brody, Amanda J. Drake, Kathryn L. Evans, Rahul Gondalia, Sonja Kunze, Brigitte Kuhnel, Jari Lahti, Rozenn N. Lemaitre, Riccardo E. Marioni, Brenton Swenson, Jayandra Jung Himali, Hongsheng Wu, Yun Li, Allan F. McRae, Tom C. Russ, James Stewart, Zhiying Wang, Guosheng Zhang, Karl-Heinz Ladwig, Andre G. Uitterlinden, Xiuqing Guo, Annette Peters, Katri Räikkönen, John M. Starr, Melanie Waldenberger, Naomi R. Wray, Eric A. Whitsel, Nona Sotoodehnia, Sudha Seshadri, David J. Porteous, Joyce van Meurs, Thomas H. Mosley, Andrew M. McIntosh, Michael M. Mendelson, Daniel Levy, Lifang Hou, Johan G. Eriksson, Myriam Fornage, Ian J. Deary, Andrea Baccarelli, Henning Tiemeier, Najaf Amin.

**first authors contributed equally*

This chapter is published in the Journal of the American Medical Association (JAMA) Psychiatry, July 2018.

The supplemental material for this paper is available at: archpsyc.jamanetwork.com/article.aspx?doi=10.1001/jamapsychiatry.2018.1725

ABSTRACT

Importance: Depressive disorders arise from a combination of genetic and environmental risk factors, however, the pathophysiology and underlying molecular events leading to depression remain elusive. Epigenetic disruption provides a plausible mechanism through which gene-environment interactions lead to depression. Large-scale epigenome-wide studies on depression are missing, hampering the identification of potentially modifiable biomarkers. *Objective*: To identify robust epigenetic mechanisms underlying depression in middle-aged and elderly persons using DNA methylation in blood. *Design*: We performed the first cross-ethnic meta-analysis of epigenome-wide association studies (EWAS) within the framework of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. *Setting*: Discovery EWAS was performed in nine population-based cohorts. Results of the EWAS from all cohorts were pooled using sample-size weighted meta-analysis. Replication of the top epigenetic sites from the discovery stage was performed in two independent population-based cohorts. *Participants*: The discovery sample included 7,948 individuals of European origin and the replication sample included 3,308 individuals of African-American and European origin. Only participants that were assessed for both depressive symptoms and whole blood DNA methylation were included in the study. *Outcome*: Whole blood DNA methylation levels were assayed with Illumina-Infinium Human Methylation 450K BeadChip and depressive symptoms were assessed by questionnaire. *Results*: The discovery cohorts consisted of 7,948 individuals (48% female) with a mean age of 65.4 (SD=5.8) years. The replication cohort consisted of 3,308 individuals (74% female) with a mean age of 60.3 (SD=6.4) years. The EWAS identified methylation of three CpG sites including cg04987734 (*p-value*=1.57×10^{-8}, *n*=11256, *CDC42BPB* gene), cg12325605 (*p-value*=5.24×10^{-9}, *n*=11256, *ARHGEF3* gene) and an intergenic CpG site cg14023999 (*p-value*=5.99×10^{-8}, *n*=11256, chromosome=15q26.1) significantly associated with increased depressive symptoms. The predicted expression of *CDC42BPB* and *ARHGEF3* was significantly associated with major depression in brain and fibroblasts, respectively. *Conclusion*: We report the first robustly associated methylated sites for depressive symptoms. All three findings point towards axon guidance as the common disrupted pathway in depression. Our findings provide new insights into the molecular mechanisms underlying the complex pathophysiology of depression. Further research is warranted to determine the utility of these findings as biomarkers of depression and evaluate any potential role in the pathophysiology of depression and their downstream clinical effects.

INTRODUCTION

Depression is one of the most common mental health disorders that is projected to play a leading role in disease burden by the year 2030.[1] In later life, depression is associated with disability, increased mortality, dementia and poor outcomes from physical illness.[2] Further, more people aged over 65 years commit suicide than in any other age group, and most have major depression.[3] Limited understanding of the molecular mechanisms underlying depression is a major bottleneck in the development of innovative treatment, prognostic markers, and prevention strategies.

Studying depression is challenging, as it is a heterogeneous disorder with a multifactorial etiology.[4] This heterogeneity increases with age as the incidence of chronic diseases and disability rises. The contribution of genetics to the risk of depression is moderate with heritability estimates ranging from 40 to 50%[5] and modest (18%) in the elderly.[6] Genome-wide association studies (GWAS) have recently identified numerous rare and common genetic variants associated with depression and related traits.[7-10] However, genetic variation alone does not completely explain an individual's risk for developing depression. Among environmental factors, adverse life-events and stress are major risk factors for depression.[11] Converging evidence from animal and human studies suggest that psychosocial stressors trigger depression onset by inducing elevations in pro-inflammatory cytokines.[12] These psychosocial stressors are also known to influence epigenetic mechanisms, such as DNA methylation[13] that can drive sustained changes in gene expression.[14] The high contribution of environmental factors to depression in the elderly makes DNA methylation an interesting candidate mechanism for studies of the molecular basis of late-life depression.

DNA methylation may be global or tissue-specific.[15] Tissues likely to be involved in complex psychiatric disorders, such as brain, are not directly accessible from living patients. The, use of post-mortem brain tissue to study DNA methylation is a possible solution, although obtaining a sufficient sample size is challenging.[16] To study differential DNA methylation associated with mental health symptoms on a large scale, peripheral tissues such as blood constitutes a useful proxy for detecting trans-tissue changes and the most appropriate tissue for biomarkers.[16,17] Moderate correlation has been demonstrated between blood and brain tissues at non-tissue specific regulatory regions across the methylome.[18] To date, several studies have assessed the correlation between depression and blood DNA methylation.[19,20] However, these studies have been limited to a small number of DNA methylation sites (CpGs) and/or small samples. For instance, the largest published epigenome-wide association study (EWAS) assessed brain DNA methylation in 76 cases persons who

died during a depressive episode and 45 controls.[21] Moreover, these studies compared depressed cases with healthy controls. Depression represents an arbitrarily selected extreme of the continuum of varying severity and duration,[22] whereas a broad phenotype approach can be more representative for the general population. In a large study consisting of 252,503 individuals from 68 countries showed that sub-threshold depressive disorders produce significant decrements in health and do not qualitatively differ from full-blown episodes of depression.[23] A meta-analysis in individuals aged over 55 found two to three times higher prevalence of sub-threshold depressive symptomology than major depression.[24] Use of rating scales have therefore been recommended for the assessment of depressive problems in the elderly.[2]

In the current study, we performed EWAS of depressive symptoms using whole blood samples of 7,948 individuals of European ethnicity from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. We replicated our findings in 3,308 individuals of African-American and European ancestry. Finally, we used publicly available expression quantitative methylation (eQTM) loci and expression quantitative loci (eQTL) databases to identify the downstream effects of the associated methylation loci.

MATERIALS AND METHODS

Study population

The study sample for the discovery analysis included a total of 7,948 participants of European ancestry from nine population-based cohorts of the CHARGE consortium (**Table 1**): Cardiovascular Health Study (CHS),[25] Framingham Heart Study (FHS),[26] Helsinki Birth Cohort Study (HBCS),[27] Cooperative Health Research in the Augsburg Region (KORA) study,[28] two sub-cohorts from Lothian Birth-Cohort born in 1921 (LBC1921)[29] and 1936 (LBC1936),[30] two sub-cohorts from Rotterdam Study (RS-III and RS-BIOS)[31] and Generation Scotland: Scottish Family Health Study (GS) study.[32] These cohorts included community dwelling individuals, who were not selected based on disease status. Informed consent was obtained from all participants. The same cohorts have been successfully used to identify differentially methylated sites associated with cognitive traits,[33] inflammation[34] and smoking.[35] The protocol for each study was approved by the institutional review board of each institution.

The replication sample included 3,308 participants, largely of African American origin from the Atherosclerosis Risk in Communities Study (ARIC)[36] and European origin from the Women's Health Initiative - Epigenetic Mechanisms of PM-Mediated

Cardiovascular disease (WHI-EMPC) that joined the consortium later for the replication phase of the study.[37] Detailed information for each cohort is provided in the **Supplementary Text**.

Depressive symptoms assessment

Depressive symptoms were measured using self-reported questionnaires or structured interview performed by a trained researcher, psychologist, or psychiatrist at the same time point when blood samples were obtained for DNA methylation quantification (**Table 1**). Four cohorts (FHS, HBCS, RS-III, and RS-BIOS) assessed depressive symptoms using the 20-item Centre for Epidemiologic Studies Depression (CES-D) scale,[38] while CHS used the 10-item CES-D scale. Participants could score from zero to 60 (or 30 for CHS) points, where higher scores suggest more depressive symptoms. WHI-EMPC used a cohort specific CES-D/DIS screening instrument, which is described in detail in the **Supplementary Text**. The LBC1921 and LBC1936 assessed self-reported depressive symptoms using the Hospital Anxiety and Depression Scale-depression subscale (HADS-D),[39] which consists of seven items. Participants could score from zero to 21. The KORA study used the self-administered Patient Health Questionnaire (PHQ-9)[40] representing a depression module that scores each of the nine Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) criteria for depression from zero to three. The GS study assessed life-time history of depression using the Structured Clinical Interview for DSM-IV Disorders (SCID).[41] The ARIC study assessed depressive symptoms using the 21-item Maastricht Questionnaire (21-MQ). In all cohorts, depressive symptoms were analyzed as continuous variable except for GS, which studied depression status as binary trait.

DNA methylation sample and measurement

In all cohorts, DNA was extracted from whole blood and methylation levels were assessed using the Illumina-Infinium Human Methylation 450K BeadChip (Illumina Inc., San Diego, CA, USA) using standard manufacturer's protocols. The 450K array includes more than 450,000 CpGs and is enriched for genic regions, covering 99% of all genes. DNA methylation data pre-processing, including quality control (QC) and normalization, was conducted per cohort using study-specific methods. In all cohorts, DNA methylation levels were quantified as β-values, which range from zero to one, and indicate the proportion of DNA strands in a sample methylated at a specific CpG. Detailed information about cohort specific DNA extraction, bisulfite conversion, DNA methylation profiling, normalization and QC is described in detail in the **Supplementary Text**.

Statistical analysis

Epigenome-wide association analysis

In all cohorts, the association between depressive symptoms and CpG sites was assessed using linear regression analysis in the R software. In the regression analysis, DNA methylation β-value at each CpG site was specified as the dependent variable and the depressive symptoms/depression as the predictor of interest. Association analysis was adjusted for age,[42] sex,[43] smoking[35] (assessed at the time of blood sampling for methylation), methylation batch effects, white blood cell composition (imputed or directly measured), principal components estimated using genome-wide genotype data to control for population stratification and familial relationships when required. Cohort specific details of these analyses are provided in the **Supplementary Text**. Further, sensitivity analysis was performed by adjusting the initial model for antidepressant medication use at the time of sample collection.

To pool the results from independent studies we performed sample-size weighted meta-analysis in METAL.[44] We chose the 'sample-size weighted' method because of the differences in the measurement scales of depressive symptoms across studies. A drawback of using sample-size weighted method is that no pooled effect estimates are generated. To obtain pooled effect estimates we additionally performed inverse-variance weighted meta-analysis for the top sites in cohorts that used CES-D 20 item scale for the assessment of depressive symptoms. CpG sites missing in more than three of the participating cohorts were removed. In total, 484,516 probes were tested for association, giving a Bonferroni-corrected genome-wide significance threshold of $0.05/484{,}516 = 1.03\times10^{-7}$. All CpG sites suggestive of association (*p-value* $\leq 10^{-5}$) were tested for association in the independent replication cohorts using the same model as used in the discovery EWAS. Finally, a sample size weighted meta-analysis was performed for all cohorts included in the discovery and replication phases in METAL. To evaluate the contribution of each study to the association results we generated posterior probabilities of the effects in each study (M-values) using the METASOFT package.[45] M-value and Forest plots for z-scores were generated using custom-made scripts in R. For annotating CpG sites we used the annotation provided by Illumina and the UCSC database (GRCh37/hg19).

Gene expression analyses

To evaluate the downstream effects of the three identified CpG sites in blood we used the BIOS database to search for eQTM.[46] To evaluate whether the expression of the genes associated with or harbored the significant methylation site is associated with major depression (also smoking and inflammation to check specificity) we

used the MetaXcan package.[47,48] MetaXcan associates the expression of the genes with the phenotype by integrating functional data generated by large-scale efforts, e.g., Genotype-Tissue Expression (GTEx) with that of the GWAS. MetaXcan is trained on transcriptome models in 44 human tissues from GTEx and is able to estimate their tissue-specific effect on phenotypes from GWAS. We used the GTEx-V6p-HapMap-2016-09-08 database and the publicly available GWAS datasets of major depression,[49] and C-reactive protein[50] and smoking,[51] which represent important potential confounders in the present study.

Causal inference analysis

To help infer causal relationships, we studied the *cis*-SNPs identified by the BIOS consortium[46] as instrumental variables for the CpG sites as proposed by Smith et al.[52] We checked the association of these cis-SNPs with depression, smoking and inflammation in the published GWAS of these traits. Similarly, we checked whether the single nucleotide polymorphisms (SNPs) associated with inflammation (CRP levels),[50] smoking[51] and depression[7] were associated with the identified CPG sites using the BIOS consortium database. We chose smoking and inflammation as these are highly correlated with both depression and DNA methylation and thus could influence the relationship between depression and DNA methylation.

RESULTS

The mean age in the discovery cohorts ranged from 52.4 years (SD=8.1) in GS to 79.1 years (SD=0.57) in LBC1921. Forty-eight percent of the total discovery sample were female. The replication cohort comprised 74% women and had an average age of 60.3 years (SD=6.4) (**Table 1**).

Epigenome-wide association analysis

In the meta-analysis of depressive symptoms of European ancestry, we identified one CpG site on chromosome 14q32.32 (cg04987734, *CDC42BPB*, *p-value*=4.93×10^{-8}, *n*=7948) that passed the Bonferroni threshold for significance (**Table 2, eFigure 1**). Further, suggestive association was observed at 19 additional CpG sites (**Table 2**).

Adjusting for anti-depressive medication use did not meaningfully change the results (**eTable 1**). No inflation in the test statistic was observed (Lambda=1.03, **eFigure 2**). We tested all 20 CpG sites for association in the replication sample. The top CpG site from the discovery (cg04987734) showed nominal association (*p-value*<0.05, *n*=3308) with depressive symptoms in the validation data set (**Table 2**).

Table 1. Descriptive statistics of the discovery and replication cohorts.

Study	Ethnicity	N	Female (%)	Mean Age (SD)	Current smokers (%)	Depressive symptoms	Antidepressant medication use (%)
Discovery (N = 7948)							
CHS	European	323	194 (60.1)	75.6 (5.2)	173 (53.6)	CESD[1] (10 item)	19 (5.9)
FHS	European	2722	1508 (53.6)	58.5 (11.6)	948 (34.8)	CESD[2] (20 item)	251 (16.1)
HBCS	European	122	0 (0)	65.2 (2.7)	24 (19.7)	CESD[2] (20 item)	11 (9.0)
KORA	European	1727	882 (51.1)	61.0 (8.9)	250 (14.5)	PHQ-9[3]	82 (4.7)
LBC 1921	European	432	261 (60.4)	79.1 (0.6)	194 (44.9)	HADS[4]	15 (3.5)
LBC 1936	European	916	452 (49.3)	69.6 (0.8)	504 (55)	HADS[4]	30 (3.3)
RS III	European	722	391 (54.2)	59.8 (8.1)	167 (23.1)	CESD[2] (20 item)	38 (5.3)
RS BIOS	European	757	319 (42.1)	67.6 (5.9)	78 (10.3)	CESD[2] (20 item)	51 (6.7)
GS[a]	European	227	151 (64.5)	52.4 (8.1)	46 (19.7)	SCID[5]	44 (18.8)
Total		7948	4158 (48.4)	65.4 (5.8)	2384 (30.6)	-	541 (8.1)
Replication (N = 3308)							
ARIC	African	2297	1445 (63)	56.1 (5.7)	584 (25.4)	21-MQ[6]	74 (3.3)
WHI-EMPC	European	1011	1011 (100)	64.6 (7.1)	509 (50.3)	CES-D/DIS[7]	61 (6.0)
Total		3308	2456 (74.2)	60.3 (6.4)	1093 (37.9)	-	135 (4.7)

Characteristics are depicted as mean (SD), unless otherwise specified. *CHC* Cardiovascular health cohort, *FHS* Framingham Heart Study, *HBCS* Helsinki Birth Cohort Study, *KORA* Cooperative Health Research in the Augsburg Region, *LBC* Lothian Birth Cohort, *RS* Rotterdam Study, *GS* Generation Scotland Study, and [a] CASE-CONTROL STUDY, *ARIC* Atherosclerosis Risk in Communities Study and *WHI-EMPC* the Women's Health Initiative - Epigenetic Mechanisms of PM-Mediated Cardiovascular disease; "(in brackets we state number of item of the questionnaires)". 1 Irwin, M. et al. 1999. [2] Radloff, LS. et al. 1977. [3] Kroenke, K. et al. 2001. [4] Zigmond, AS. et al. 1983. [5] First, MB. et al. 1996. [6] Wattanakit,K. et al. 2005. [7] Burnam, MA. et al. 1988.

In addition, significant association of a CpG site (cg12325605; *p-value*=9.17×10^{-05}, *n*=3308, **Table 2**) annotated to the *ARHGEF3* gene with depressive symptoms was observed in the replication sample.

Meta-analysis of discovery and replication cohorts showed a significant association of both cg04987734 (*p-value*=1.57×10$^{-08,}$ *n*=11256) and cg12325605 (*p-value*=5.24×10^{-09}, *n*=11256) with depressive symptoms (**Table 2; Figures 1 and 2**).

Also, an additional intergenic CpG site (cg14023999; *p-value*=5.99×10^{-08}, *n*=11256) at chromosome 15q26.1 locus showed genome-wide significant association with depressive symptoms (**eTable 2, eFigures 3 and 4**). The independent contributions of each cohort to the association signals of the three CpG are depicted in **eFigure 5** and also provided in **eTable 3**. For all three CpG sites the association signals were not driven by a single cohort but appeared to be linearly related to the sample size, sug-

Table 2. Top DNA methylation sites associated with depressive symptoms in the discovery EWAS.

				Discovery (N=7948)	Replication (N=3308)	Meta-analysis (N=11256)
CpG site ID	**Chr**	**Location**	**Gene symbol**	**P-value**	**P-value**	**P-value**
cg04987734	14	103415873	*CDC42BPB*	**4.93×10^{-8}**	4.82×10^{-02}	**1.57×10^{-08}**
cg07012687	17	80195180	*SLC16A3*	3.47×10^{-7}	1.58×10^{-01}	4.45×10^{-06}
cg08796240	16	70733832	*VAC14*	7.43×10^{-7}	2.56×10^{-01}	1.80×10^{-06}
cg06096336	2	231989800	*PSMD1; HTR2B*	8.06×10^{-7}	3.01×10^{-01}	2.51×10^{-06}
cg16745930	10	100220809	*HPSE2*	1.34×10^{-6}	4.01×10^{-01}	6.26×10^{-06}
cg09849319	5	1494983	*LPCATI1*	1.81×10^{-6}	4.64×10^{-01}	1.04×10^{-04}
cg17237086	22	40814966	*MKL1*	3.44×10^{-6}	2.51×10^{-01}	6.10×10^{-06}
cg03985718	2	105924245	*TGFBRAP1*	3.61×10^{-6}	8.54×10^{-01}	6.53×10^{-05}
cg21098005	20	44538605	*PLTP*	4.36×10^{-6}	9.60×10^{-01}	1.01×10^{-04}
cg16466652	19	6271960	*MLLT1*	4.39×10^{-6}	3.97×10^{-01}	1.57×10^{-05}
cg07884764	11	64107517	*CCDC88B*	5.03×10^{-6}	9.99×10^{-01}	1.25×10^{-04}
cg01541347	7	4729920	*FOXK1*	5.64×10^{-6}	3.77×10^{-01}	8.46×10^{-04}
cg02341197	21	34185927	*C21or62*	5.84×10^{-6}	2.02×10^{-01}	6.80×10^{-06}
cg01947751	3	196728969	-	6.23×10^{-6}	6.63×10^{-01}	3.68×10^{-04}
cg13747876	17	80195402	*SLC16A3*	6.32×10^{-6}	1.04×10^{-01}	2.93×10^{-06}
cg12764201	1	105101123	*CORT; APITD1*	7.15×10^{-6}	7.20×10^{-01}	7.29×10^{-05}
cg08295111	5	133866097	*PHF15*	7.87×10^{-6}	5.76×10^{-01}	5.64×10^{-04}
cg18030453	3	45506216	*LARS2*	9.16×10^{-6}	3.87×10^{-03}	1.20×10^{-07}
cg12325605	3	56810151	*ARHGEF3*	9.62×10^{-6}	**9.17×10^{-05}**	**5.24×10^{-09}**
cg23282441	10	73533927	*C10orf54; CDH23*	9.69×10^{-6}	1.77×10^{-01}	8.63×10^{-06}

gesting stronger association in larger studies (**eFigure 5**). Pooled effect estimates in cohorts that used CES-D scale suggest that a 1-unit increase in CES-D score increases methylation by 0.05% at cg04987734, 0.04% at cg12325605, and 0.03% at cg14023999.

Gene expression analyses

Cg04987734 was significantly associated with increased expression of CDC42BPB gene (FDR *p-value*=7.7×10^{-04}, *n*=2101) and cg14023999 was significantly associated with decreased expression of SEMA4B (FDR *p-value*=4.7×10^{-03}, *n*=2101) in blood (**eTable 4**). No significantly associated gene expression probes were identified for cg12325605 in blood. Further, the predicted expression of *CDC42BPB* gene in the brain (basal ganglia) (effect=0.14, *p-value*=2.7×10^{-03}) and of *ARHGEF3* in fibroblasts (effect=-0.48, *p-value*=9.8×10^{-04}) was associated with major depression (**eTable 5**). No association was observed with either smoking or inflammation.

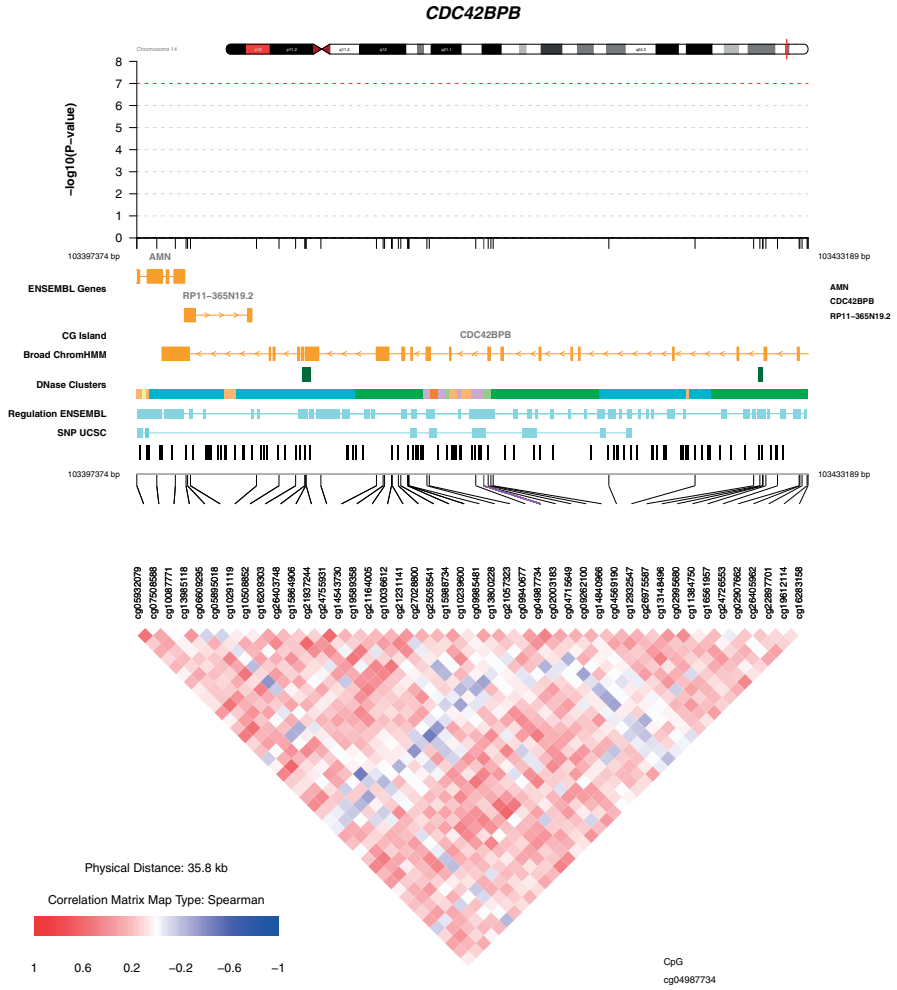

Figure 1: Regional association plot for the top CpG site cg04987734. The horizontal axis depicts the position in base pair (hg19) for the entire *CDC42BPB* gene region. The Vertical axis indicates the strength of association in terms of negative logarithm of the association *p*-value. Each circle represents CpG site. Red dashed line indicates the genome-wide significance threshold. Below the horizontal axis the figure shows the regulatory information and correlation matrix of other CpG sites in the region with the top hit. Color intensity marks the strength of the correlation and color the direction of the correlation. Figure is made using web-based plotting tool and R-based package "CoMET" (http://comet.epigen.kcl.ac.uk:3838/coMET1/).

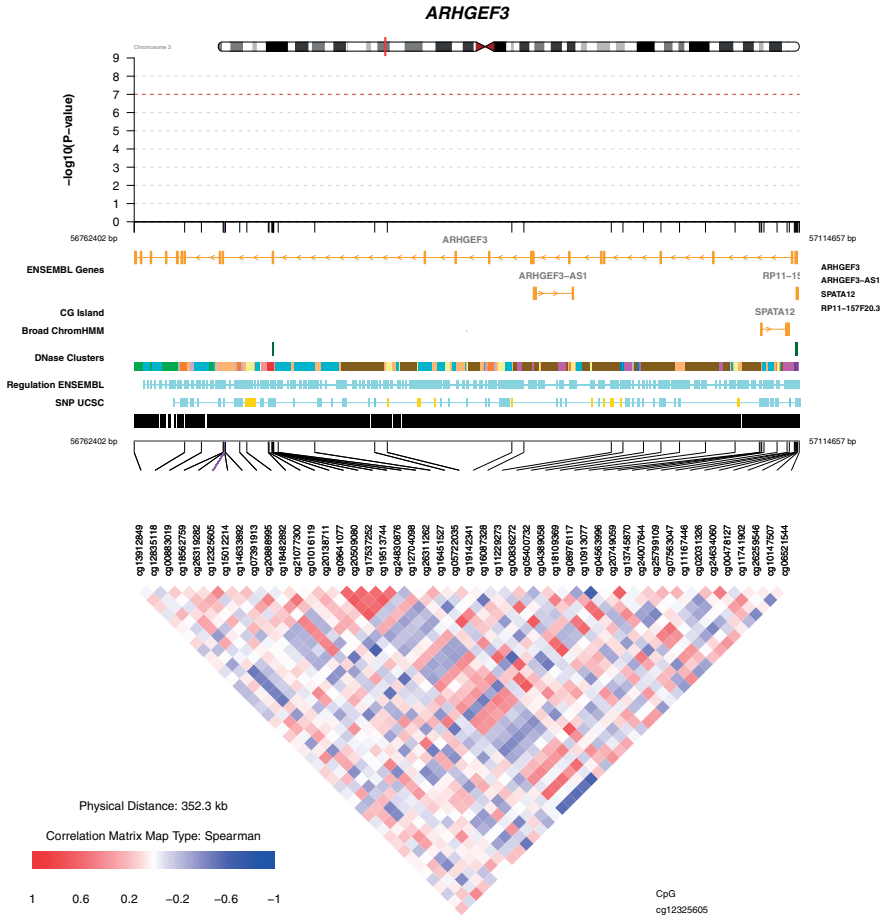

Figure 2: Regional association plot for the top CpG site cg12325605. The horizontal axis depicts the position in base pair (hg19) for the entire *ARHGEF3* gene region. The Vertical axis indicates the strength of association in terms of negative logarithm of the association *p*-value. Each circle represents CpG site. Red dashed line indicates the genome-wide significance threshold. Below the horizontal axis the figure shows the regulatory information and correlation matrix of other CpG sites in the region with the top hit. Color intensity marks the strength of the correlation and color the direction of the correlation. Figure is made using web-based plotting tool and R-based package "CoMET" (http://comet.epigen.kcl.ac.uk:3838/coMET1/).

Blood and brain correlation

We checked the correlation between methylation in blood and various brain regions at the three identified sites using a web-based tools, BECon[18] and a blood brain DNA methylation comparison tool (http://epigenetics.essex.ac.uk/bloodbrain/). BECon showed strong correlation between blood and brain DNA methylation, e.g. methylation at cg04987734 in the CDC42BPB gene was highly correlated (r=0.81) between blood and the Brodmann area 7 that spans the medial and lateral walls of the parietal cortex (**eFigure 6**). Methylation at the other two sites was negatively correlated with methylation in the Brodmann area 10 than spans anterior prefrontal cortex (cg12325605, r=-0.39; cg14023999, r=-0.42) suggesting strong but reverse methylation patterns in blood and brain (**eFigures 7 and 8**). However, the blood brain DNA methylation comparison tool that compares DNA methylation between blood and prefrontal cortex, entorhinal cortex, superior temporal gyrus and cerebellum, showed only modest correlations. For instance, methylation in blood at cg04987734 showed the strongest correlation with methylation in superior temporal gyrus (r = 0.18; http://epigenetics.essex.ac.uk/bloodbrain/?probe nameg=cg04987734), while methylation in blood at cg12325605 (http://epigenetics .essex.ac.uk/bloodbrain/?probenameg=cg12325605) and cg14023999 (http:// epigenetics.essex.ac.uk/bloodbrain/?probenameg=cg14023999) showed strongest correlation with methylation in cerebellum (r = 0.16 and 0.19 respectively). Nevertheless, the findings from the two databases suggest some degree of correlation between methylation in blood and methylation in brain for the three identified CpG sites.

Causal inference

In the BIOS database we identified two cis-SNPs for cg04987734 and 4 cis-SNPs for cg12325605 (**eTable 6**) and none for cg14023999. We took the most significant cis-SNP as the proxy for the CpG sites if available. For cg04987734 we used rs751837 as a proxy and for cg12325605 we used rs3821412 as a proxy (top cis-SNP rs9880418 was not available in the GWAS of depression, smoking or inflammation). Rs751837 was suggestively associated with major depression (p-value=0.07; albeit in opposite direction) (**eTable 7**). Rs3821412 was not associated with any of the three tested phenotypes. None of the SNPs associated with depression, inflammation or smoking was associated with any of the three CpG sites.

DISCUSSION

In this large-scale EWAS of depressive symptoms, we identified methylation at three CpG sites (cg04987734, cg12325605 and cg14023999) associated with depressive symptoms in the middle-aged and elderly persons. Cg04987734 is annotated to the *CDC42BPB* gene, cg12325605 to the *ARHGEF3* gene, and cg14023999 lies in an intergenic region on chromosome 15q26.1 locus. The predicted expression of *CDC42BPB* and *ARHGEF3* genes associate with major depression in brain and fibroblasts respectively.

CDC42BPB (CDC42 Binding Protein Kinase Beta) encodes a member of the serine/threonine protein kinase family, which is an important downstream effector of CDC42 and plays a role in the regulation of cytoskeleton reorganization, cell migration and regulation of neurite outgrowth.[53] CDC42BPB is highly expressed in the brain https://www.proteinatlas.org/ENSG00000198752-CDC42BPB/tissue. Hyper-methylation of cg04987734 has been associated with increased expression of CDC42BPB in blood.[46] Interestingly, methylation levels at this CpG site (cg04987734) in *CDC42BPB* gene were also previously associated with C-reactive protein (CRP) levels in blood;[34] and smoking.[35] In our study, however, we adjusted for smoking in the regression model; therefore, the association between depression and DNA methylation of this CpG site may be independent of smoking habits. Also, our causal inference analyses provide no support for the possibility that smoking, or inflammation explained the observed association with depressive symptoms nor the predicted expression of the gene showed an association with smoking or inflammation.

ARHGEF3 encodes for Rho Guanine Nucleotide Exchange Factor 3 protein. The gene is highly expressed (https://www.proteinatlas.org/ENSG00000163947-ARHGEF3/tissue) in adrenal glands, brain and uterus. Both *ARHGEF3* and *CDC42BPB* are co-expressed with several members of the Rho subfamily (RHOA, RHOB and RHOC; **eFigures 9 and 10**) of the Rho GTPase family that also includes CDC42.[54] The Rho family of GTPases is a family of small signaling G proteins involved in p75 neurotrophin receptor (p75NTR)-mediated signaling[55] and semaphorin signaling pathways.[56] P75NTR is a transmembrane receptor for neurotrophic factors of the neurotrophin family, which includes the brain-derived neurotrophic factor (BDNF).[57] P75NTR is widely expressed in the developing central and peripheral nervous system during the period of synaptogenesis and developmental cell death.[58] Both p75NTR and semaphorins are implicated in axon guidance.[59,60] In this context, the third associated CpG site cg14023999 that lies in an intergenic region on chromosome 15q26.1 is also interesting. Cg14023999 is associated with decreased expression of *SEMA4B* gene in blood. *SEMA4B* encodes for Semaphorin 4B protein. Sema4B is

believed to function through a direct interaction with post-synaptic density protein PSD-95[61] to promote synapse maturation.[61-63] The knock-down of Sema4B causes a decrease in GABAergic synapse number[62] suggesting a role in the assembly of excitatory and inhibitory postsynaptic specializations.[63] Previously cg14023999 was found to be significantly correlated with Parkinson's disease[64] and significant association of a CpG site in *SEMA4B* was observed in individuals with schizophrenia carrying the 22q11.2 deletion.[65] These findings point towards a functional of *SEMA4B* in neuro-psychiatric disorders. When comparing our findings with that of the previous EWAS of depression, we did not find an overlap. These studies were small (<100 individuals) and did not report reproducible results.[20]

To summarize, we report the first EWAS of depressive symptoms. We identified and replicated association of two methylation sites in the genome with depressive symptoms. A third site was identified in the meta-analysis of discovery and replication cohorts, which requires further replication. All three findings point towards axon guidance as the common disrupted pathway in depression (http://www.genome.jp/kegg/pathway/hsa/hsa 04360.html). Our findings provide new insights into the molecular mechanisms underlying the complex pathophysiology of depression.

Strengths and limitations of the study

This is the largest epigenome-wide study of depressive symptoms reported to date. Our major strength is the sample size that enabled detection of a replicable epigenome-wide significant locus, which suggests that in blood, DNA methylation signatures associated with depression may be subtle and will require large samples to be detected. Using peripheral blood tissue for DNA methylation profiling is a limitation of this study, as DNA methylation is known to be tissue specific.[66] While peripheral blood is not considered to be the most relevant tissue for the pathophysiology of depression, some sites show correlated methylation profiles between-tissues.[15,66] The three sites identified in our study show some degree of correlation between methylation in blood and various brain regions. Second, while replication in African-American samples suggests that some depressive symptoms related differences in DNA methylation may be similar across ethnicities,[67-69] it may also have resulted in false negatives due to different genetic background. Third, in these analyses we mostly used quantitative measures of depressive symptoms. Quantitative endo-phenotypes provide powerful alternatives for several complex outcomes, for example, hypertension.[70] This is likely to be especially true for a trait such as depressive symptoms, for which the severity and duration of illness can be highly heterogeneous.[22] Genome-wide studies of depressive traits, using quantita-

tive endo-phenotypes, have been suggested to improve statistical power.[22] However, the use of different phenotypic measures by different cohorts means that there may be some loss of statistical power due to the heterogeneity in the phenotype assessment. Nevertheless, the top three sites in our study were robustly associated with depressive symptoms independent of the depressive symptom measure used. Fourth, although we adjusted for potential confounders, the possibility of residual confounding cannot be excluded. Antidepressant medication indicates treated depression but itself may result in epigenetic modifications involved in depression pathophysiology.[71] Antidepressants can thus mediate or confound the relation between DNA methylation and depression. However, in sensitivity analysis additionally adjusted for antidepressant medication, our results did not change. Fifth, most cohorts included in this EWAS are cohorts of elderly persons. The aetiology of depression is more heterogeneous in elderly people than in younger and often hidden behind somatic symptoms, either because of somatization of the disorder or because of accentuation of symptoms of concomitant illness.[72] This may affect the generalizability of the results to younger populations. Finally, we made an attempt to disentangle cause and consequence using SNPs associated with the identified CpG sites and depression, inflammation and smoking as instrumental variables. The results did not support a causative role yet the association of the predicted gene expression of CDC42BPB in brain and ARHGEF3 in fibroblasts with major depression does suggest a possible causal role of the regulatory effects of these genes.[47]

REFERENCES

1. Lepine JP, Briley M. The increasing burden of depression. *Neuropsychiatr Dis Treat.* 2011; 7(Suppl 1):3-7.
2. Rodda J, Walker Z, Carter J. Depression in older adults. *BMJ.* 2011;343:d5219.
3. Conwell Y, Duberstein PR, Cox C, Herrmann JH, Forbes NT, Caine ED. Relationships of age and axis I diagnoses in victims of completed suicide: a psychological autopsy study. *Am J Psychiatry.* 1996;153(8):1001-1008.
4. Hidaka BH. Depression as a disease of modernity: explanations for increasing prevalence. *J Affect Disord.* 2012;140(3):205-214.
5. Lohoff FW. Overview of the genetics of major depressive disorder. *Curr Psychiatry Rep.* 2010; 12(6):539-546.
6. Gatz M, Pedersen NL, Plomin R, Nesselroade JR, McClearn GE. Importance of shared genes and shared environments for symptoms of depression in older adults. *J Abnorm Psychol.* 1992;101(4):701-708.
7. Wray NR, Sullivan PF. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *bioRxiv.* 2017.
8. CONVERGE. consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature.* 2015;523:588–591.
9. Hek K, Demirkan A, Lahti J, et al. A genome-wide association study of depressive symptoms. *Biol Psychiatry.* 2013;73(7):667-678.
10. Direk N, Williams S, Smith JA, et al. An Analysis of Two Genome-wide Association Meta-analyses Identifies a New Locus for Broad Depression Phenotype. *Biol Psychiatry.* 2017; 82(5):322-329.
11. Agid O, Kohn Y, Lerer B. Environmental stress and psychiatric illness. *Biomed Pharmacother.* 2000;54(3):135-141.
12. Berk M, Williams LJ, Jacka FN, et al. So depression is an inflammatory disease, but where does the inflammation come from? *BMC Med.* 2013;11:200.
13. Dahl C, Guldberg P. DNA methylation analysis techniques. *Biogerontology.* 2003;4(4):233-250.
14. Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. Metastable epialleles in mammals. *Trends Genet.* 2002;18(7):348-351.
15. Lokk K, Modhukur V, Rajashekar B, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol.* 2014;15(4):r54.
16. Byrne EM, Carrillo-Roa T, Henders AK, et al. Monozygotic twins affected with major depressive disorder have greater variance in methylation than their unaffected co-twin. *Transl Psychiatry.* 2013;3:e269.
17. Latham KE. Stage-specific and cell type-specific aspects of genomic imprinting effects in mammals. *Differentiation.* 1995;59(5):269-282.
18. Edgar RD, Jones MJ, Meaney MJ, Turecki G, Kobor MS. BECon: A tool for interpreting DNA methylation findings from blood in the context of brain. *bioRxiv.* 2017.
19. Bakusic J, Schaufeli W, Claes S, Godderis L. Stress, burnout and depression: A systematic review on DNA methylation mechanisms. *J Psychosom Res.* 2017;92:34-44.
20. Januar V, Saffery R, Ryan J. Epigenetics and depressive disorders: a review of current progress and future directions. *Int J Epidemiol.* 2015;44(4):1364-1387.
21. Nagy C, Suderman M, Yang J, et al. Astrocytic abnormalities and global DNA methylation patterns in depression and suicide. *Mol Psychiatry.* 2015;20(3):320-328.

22. van der Sluis S, Posthuma D, Nivard MG, Verhage M, Dolan CV. Power in GWAS: lifting the curse of the clinical cut-off. *Mol Psychiatry.* 2013;18(1):2-3.
23. Ayuso-Mateos JL, Nuevo R, Verdes E, Naidoo N, Chatterji S. From depressive symptoms to depressive disorders: the relevance of thresholds. *Br J Psychiatry.* 2010;196(5):365-371.
24. Meeks TW, Vahia IV, Lavretsky H, Kulkarni G, Jeste DV. A tune in "a minor" can "b major": a review of epidemiology, illness course, and public health implications of subthreshold depression in older adults. *J Affect Disord.* 2011;129(1-3):126-142.
25. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol.* 1991;1(3):263-276.
26. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol.* 1979;110(3):281-290.
27. Eriksson JG, Forsen T, Tuomilehto J, Osmond C, Barker DJ. Early growth and coronary heart disease in later life: longitudinal study. *BMJ.* 2001;322(7292):949-953.
28. Wichmann HE, Gieger C, Illig T, Group MKS. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen.* 2005;67 Suppl 1: S26-30.
29. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol.* 2012;41(6):1576-1584.
30. Deary IJ, Gow AJ, Taylor MD, et al. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr.* 2007;7:28.
31. Ikram MA, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol.* 2017;32(9):807-850.
32. Smith BH, Campbell A, Linksted P, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol.* 2013;42(3):689-700.
33. Marioni RE, McRae AF, Bressler J, et al. Meta-analysis of epigenome-wide association studies of cognitive abilities. *Mol Psychiatry.* 2018.
34. Ligthart S, Marzi C, Aslibekyan S, et al. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.* 2016;17(1):255.
35. Joehanes R, Just AC, Marioni RE, et al. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet.* 2016;9(5):436-447.
36. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol.* 1989;129(4):687-702.
37. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials.* 1998;19(1):61-109.
38. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB. Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychol Aging.* 1997;12(2):277-287.
39. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;67(6):361-370.
40. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9):606-613.

41. First MB, Spitzer, Robert L, Gibbon Miriam, and Williams, Janet B.W. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P) November 2002.
42. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14(10): R115.
43. Singmann P, Shem-Tov D, Wahl S, et al. Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenetics Chromatin.* 2015;8:43.
44. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.
45. Han B, Eskin E. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet.* 2012;8(3):e1002555.
46. Bonder MJ, Luijk R, Zhernakova DV, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet.* 2017;49(1):131-138.
47. Barbeira A, Dickinson SP, Torres JM, et al. Integrating tissue specific mechanisms into GWAS summary results. *bioRxiv.* 2017.
48. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet.* 2015;47(9):1091-1098.
49. Major Depressive Disorder Working Group of the Psychiatric GC, Ripke S, Wray NR, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry.* 2013;18(4):497-511.
50. Dehghan A, Dupuis J, Barbalic M, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation.* 2011;123(7): 731-738.
51. Tobacco, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* 2010;42(5):441-447.
52. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol.* 2012;41(1):161-176.
53. Chen XQ, Tan I, Leung T, Lim L. The myotonic dystrophy kinase-related Cdc42-binding kinase is involved in the regulation of neurite outgrowth in PC12 cells. *J Biol Chem.* 1999;274(28): 19901-19905.
54. Etienne-Manneville S, Hall A. Rho GTPases in cell biology. *Nature.* 2002;420(6916):629-635.
55. DeGeer J, Lamarche-Vane N. Rho GTPases in neurodegeneration diseases. *Exp Cell Res.* 2013; 319(15):2384-2394.
56. Liu BP, Strittmatter SM. Semaphorin-mediated axonal guidance via Rho-related G proteins. *Curr Opin Cell Biol.* 2001;13(5):619-626.
57. Woo NH, Teng HK, Siao CJ, et al. Activation of p75NTR by proBDNF facilitates hippocampal long-term depression. *Nat Neurosci.* 2005;8(8):1069-1077.
58. Davies AM. Nerve growth factor synthesis and nerve growth factor receptor expression in neural development. *Int Rev Cytol.* 1991;128:109-138.
59. Fiore R, Puschel AW. The function of semaphorins during nervous system development. *Front Biosci.* 2003;8:s484-499.
60. Schecterson L, Bothwell M. An all-purpose tool for axon guidance. *Neuroscience.* 2008;1(47): Pe50.

61. Burkhardt C, Muller M, Badde A, Garner CC, Gundelfinger ED, Puschel AW. Semaphorin 4B interacts with the post-synaptic density protein PSD-95/SAP90 and is recruited to synapses through a C-terminal PDZ-binding motif. *FEBS Lett.* 2005;579(17):3821-3828.
62. Paradis S, Harrar DB, Lin Y, et al. An RNAi-based approach identifies molecules required for glutamatergic and GABAergic synapse development. *Neuron.* 2007;53(2):217-232.
63. Pasterkamp RJ, Giger RJ. Semaphorin function in neural plasticity and disease. *Curr Opin Neurobiol.* 2009;19(3):263-274.
64. Chuang YH, Paul KC, Bronstein JM, Bordelon Y, Horvath S, Ritz B. Parkinson's disease is associated with DNA methylation levels in human blood and saliva. *Genome Med.* 2017;9(1):76.
65. Starnawska A, Hansen CS, Sparso T, et al. Differential DNA methylation at birth associated with mental disorder in individuals with 22q11.2 deletion syndrome. *Transl Psychiatry.* 2017;7(8):e1221.
66. Bagot RC, Labonte B, Pena CJ, Nestler EJ. Epigenetic signaling in psychiatric disorders: stress and depression. *Dialogues Clin Neurosci.* 2014;16(3):281-295.
67. Klengel T, Pape J, Binder EB, Mehta D. The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology.* 2014;80:115-132.
68. Galanter JM, Gignoux CR, Oh SS, et al. Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *Elife.* 2017;6.
69. Kader F, Ghai M. DNA methylation-based variation between human populations. *Mol Genet Genomics.* 2017;292(1):5-35.
70. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011;478(7367):103-109.
71. Lisoway AJ, Zai CC, Tiwari AK, Kennedy JL. DNA Methylation and Clinical Response to Antidepressant Medication in Major Depressive Disorder: A Review and Recommendations. *Neurosci Lett.* 2017.
72. Gottfries CG. Is there a difference between elderly and younger patients with regard to the symptomatology and aetiology of depression? *Int Clin Psychopharm.* 1998;13:S13-S18.