

# Brain differential gene expression and blood cross-validation of a molecular signature of patients with major depressive disorder

Hugo Gomez Rueda<sup>a</sup> and Juan Bustillo<sup>a,b</sup>

**Introduction** The agreement between clinicians diagnosing major depressive disorder (MDD) is poor. The objective of this study was to identify a reproducible and robust gene expression marker capable of differentiating MDD from healthy control (HC) subjects.

**Materials and methods** Brain and blood gene expression datasets were searched, which included subjects with MDD and HC. The largest database including different areas of brain samples (GSE80655) was used to identify an initial gene expression marker. Tests of robustness and reproducibility were then implemented in 13 brain and 7 blood independent datasets. Correlations between expression in brain and blood samples were also examined. Finally, an enrichment analysis to explore the marker biological meaning was completed.

**Results** Twenty-eight genes were differentially expressed in GSE80655, of which 23 were critical to differentiate MDD from HC. The accuracy obtained using the 23 genes was 0.77 and 0.8, before and after the forward selection model, respectively. The gene marker's robustness and reproducibility were between the range of 0.46 and 0.63 in the other brain datasets and between 0.45 and 0.78 for

the blood datasets. Brain and blood expression tended to correlate in some samples. Thirteen of the 23 genes were related to stress and immune response.

**Conclusion** A 23 gene expression marker was able to distinguish subjects with MDD from HC, with adequate reproducibility and low robustness in the independent databases investigated. This gene set was similarly expressed in the brain and blood and involved genes related to stress and immune response. *Psychiatr Genet* 32: 105–115 Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc.

Psychiatric Genetics 2022, 32:105–115

**Keywords:** accuracy, brain, datasets, gene expression, immune response, major depressive disorder, reproducibility, robustness

<sup>a</sup>Department of Psychiatry and Behavioral Sciences, University of New Mexico Health Sciences Center and <sup>b</sup>Department of Neurosciences, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA

Correspondence to Hugo Gomez Rueda, MD, PhD, Department of Psychiatry and Behavioral Sciences, University of New Mexico Health Sciences Center, 2400 Tucker Ave NE MSC09 5030, Albuquerque, NM, USA  
Tel: 1+(505) 272 2223; e-mail: hgomezrueda@salud.unm.edu

Received 6 February 2021 Accepted 15 November 2021

## Introduction

Major depressive disorder (MDD) is highly prevalent ('NIMH » Major Depression', n.d.) and a leading cause of disability, affecting about 4.4% of the world population in the year 2017 (Friedrich, 2017). However, the diagnostic reliability of MDD has been questioned. According to the field trials of the DSM-5 Mood Disorders Working Group, the agreement between clinicians diagnosing MDD only achieved an intraclass kappa of 0.28 [95% confidence interval (CI): 0.20–0.35] (Regier *et al.*, 2013). This limited agreement has been also observed in other studies (Enns *et al.*, 2000; Canuto *et al.*, 2016; Behera *et al.*, 2017).

As in other psychiatric disorders, the development of biomarkers has been pursued to improve the diagnostic

reliability of MDD. Hence, many brain and peripheral features have been examined such as growth factors, cytokines, inflammatory markers, oxidative stress markers, endocrine markers, energy balance hormones, structural and functional imaging markers (Hacimusalar and Eşel, 2018). Though several biomarker studies for MDD have been performed (Spijker *et al.*, 2010; Savitz *et al.*, 2013; Sullivan *et al.*, 2013; Chang *et al.*, 2014; Liu *et al.*, 2014; Wingo *et al.*, 2015; Pantazatos *et al.*, 2017; Hacimusalar and Eşel, 2018), the results thus far have shown limitations in reproducibility and robustness (Savitz *et al.*, 2013; Sullivan *et al.*, 2013; Chang *et al.*, 2014; Pantazatos *et al.*, 2017). Robustness refers to the persistence of an effect in settings different from and outside of an experimental framework. Results' reproducibility refers to obtaining the same results from the conduct of an independent study that uses the same procedures (Goodman *et al.*, 2016). There are no currently available biomarkers for MDD or any psychiatric disorder to assist the clinician diagnostically (Hacimusalar and Eşel, 2018). Because the heritability of MDD is 37% (95% CI: 31–42%) (Sullivan *et al.*, 2000), genomic studies have been

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, [www.psychgenetics.com](http://www.psychgenetics.com).

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

at the forefront of biomarker development for depression (Gadad *et al.*, 2018). Many psychiatric genomic studies have used peripheral blood mononuclear cell (PBMC) as a practical source of genetic material. Because of the immune disturbances reported in MDD (Bahn *et al.*, 2001; Li *et al.*, 2004; Tomita *et al.*, 2004; Franz *et al.*, 2005; Strawbridge *et al.*, 2017; Gao *et al.*, 2018; Bhak *et al.*, 2019), PBMCs offer other potential advantages as a tissue source for biomarkers. However, the impermeability of the blood–brain barrier prevents the PBMCs access to the brain in normal conditions (Erdő *et al.*, 2017), questioning their relevance to disorders of the brain, like MDD. Still, it has been reported that gene expression from blood and brain are somewhat correlated in several medical conditions (Canuto *et al.*, 2016; Arion *et al.*, 2017; Ramaker *et al.*, 2017), including MDD (Tolentino and Schmidt, 2018). Furthermore, concordance between gene expression variation in postmortem brain and in PBMCs would validate any potential peripheral biomarkers (Deep-Soboslay *et al.*, 2011; Harrison, 2011; McCullumsmith and Meador-Woodruff, 2011). However, for obvious practical reasons, no biomarker studies of gene expression have jointly examined postmortem brain and PBMC in MDD. Still, the complementary examination of gene expression in brain and PBMCs from different cohorts of subjects with MDD could advance the development of genomic biomarkers.

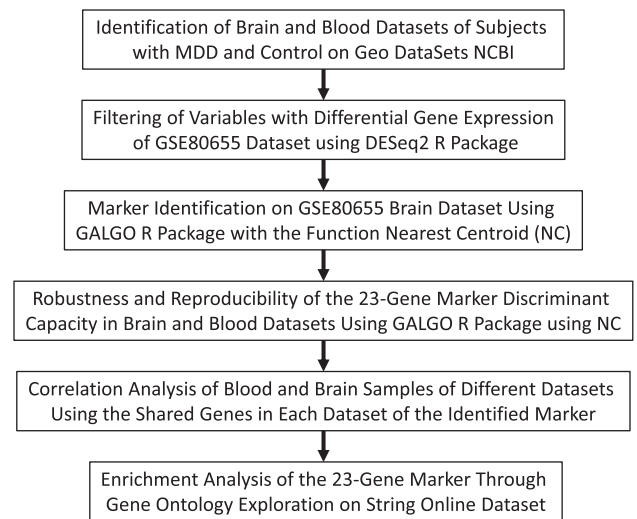
As previously stated, there is a low concordance in the diagnosis of MDD between clinicians. A biomarker could improve this concordance and the accuracy in the diagnosis of MDD. Most of the previous gene expression studies in MDD lack tests of reproducibility and robustness, and are limited to the population where the biomarker was identified. In the present study, with access to multiple datasets, we performed extensive robustness and reproducibility tests. In an effort to identify a reproducible and robust gene expression biomarker capable of differentiating MDD from healthy control (HC) subjects, we searched all the relevant online databases available up to October 2019. We started by examining protein-coding genes expressed in brain that could differentiate MDD from HC. We then tested the expression of this gene set in PBMCs in several other datasets. Subsequently, we examined the correlation of brain and blood samples of subjects with MDD and HC participants using the gene expression marker. Finally, we performed an enrichment analysis to determine the biological implications of the set of genes.

## Methods

### Identification of brain and blood datasets on Gene Expression Omnibus DataSets of NCBI

The workflow of the present study is summarized in Fig. 1. The search of databases was completed employing the terms used on Gene Expression Omnibus (GEO) DataSets (<https://www.ncbi.nlm.nih.gov/gds>) of NCBI:

Fig. 1



Summarization of material and methods.

‘MDD’ AND ‘Gene expression’, ‘MDD’ AND ‘RNA expression’, ‘Major Depressive Disorder’ AND ‘Gene expression’, ‘Major Depressive Disorder’ AND ‘RNA expression’, ‘Depression’ AND ‘Gene expression’, and ‘Depression’ AND ‘RNA expression’. The robustness and reproducibility tests were completed in datasets of brain and blood, also identified with the previously mentioned terms.

### Analysis and filtering of variables

After the selection of appropriate datasets was completed, a visual inspection of the data was performed, to confidently ascertain datasets with no missing gene expression data, meaning no empty cells in the data matrix. The platforms used to measure gene expression were microarrays or RNASeq (Table 1). The microarray datasets were normalized using quantile normalization, to control as much as possible the technical noise of the microarrays. Moreover, all the datasets included in this study were scale rank normalized with a distribution from 0 to 1, a process denominated uniformization, to address the platform differences between RNASeq and microarrays. Dataset GSE80655 was normalized with the DESeq2 R package (Love *et al.*, 2014), with the function *varianceStabilizationTransformation*.

Due to its larger sample, greater number of brain structures contained in the dataset, and use of RNASeq data, database GSE80655 was selected to first examine a molecular signature based on gene expression. This dataset was filtered for differential gene expression between patients with MDD and HCs using the function *DESeq* of the DESeq2 R Package. This procedure was completed to include in the analysis only genes with different expressions in both classes, that is, a

Table 1 Classification accuracy in datasets

Dataset ID	Platform/tissue	Total no. of probes/total no. of genes of the marker present	Subjects per class	Accuracy
Brain				
GSE80655	RNASeq/DLPFC, nAcc, anCg	23/23	HC 70, MDD 69	0.8
GSE87610	Micro/DLPFC	48/22	HC 15, MDD 19	0.63
GSE101521	RNASeq/DLPFC	23/23	HC 28, MDD 31	0.62
GSE54564	Micro/Amyg	31/20	HC 21, MDD 21	0.59
GSE44593	Micro/Amyg	27/19	HC 14, MDD 14	0.58
GSE54570	Micro/DLPFC	19/15	HC 13, MDD 13	0.58
GSE54562 + 54563	Micro/anCg	31/20	HC 35, MDD 35	0.55
GSE54575	Micro/orbitoPFC	19/15	HC 12, MDD 12	0.51
GSE53987	Micro/Hip, Stra, PFC	31/18	HC 55, MDD 50	0.49
GSE54565 + 54572	Micro/anCg	31/20	HC 28, MDD 28	0.47
GSE54567 + 54568	Micro/DLPFC	30/20	HC 29, MDD 29	0.46
Blood PBMC				
GSE38206	Micro/PBMC	38/19	HC 9, MDD 9	0.78
GSE76826	Micro/PBMC	48/21	HC 12, MDD 10	0.64
GSE39653	Micro/PBMC	29/19	HC 24, MDD 21	0.62
GSE98793	Micro/PBMC	31/18	HC 64, MDD 128	0.59
GSE19738	Micro/PBMC	25/19	HC 34, MDD 33	0.51
GSE67663	Micro/PBMC	12/9	HC 72, MDD 112	0.51
GSE52790	Micro/PBMC	20/19	HC 12, MDD 10	0.45

Amyg, amygdala; anCg, anterior cingulate; DLPFC, dorsolateral prefrontal cortex; FSM, forward selection model; HC, healthy control; Hip, hippocampus; MDD, patient with major depressive disorder; Micro, microarrays; nAcc, nucleus accumbens; OrbitoPFC, orbito prefrontal cortex; PBMC, peripheral blood mononuclear cell; PFC, prefrontal cortex; RNASeq, RNA expression measured by sequencing of RNA; Stra, striatum.

gene with high expression in HC and low expression in MDD or vice versa. This means, the classification between MDD and HC used the combined differential weighted gene expression level as a continuous variable; no cutoff of gene expression group difference was used. The final combination of genes was selected with a classification algorithm designed to improve accuracy, which is explained in more detail in the next section. The protein-coding genes with an adjusted  $P$  value in the differential gene expression less than 0.05, were selected to identify the gene expression marker. The  $P$  value of the differential gene expression was calculated with a false discovery rate method (Benjamini and Hochberg, 1995) only in protein-coding genes, to avoid type I errors in null hypothesis testing when conducting multiple comparisons. Finally, the decision of working only with protein-coding genes was determined by the lack of additional datasets with noncoding RNA to test the performance of the biomarker.

### GALGO R package for marker identification and feature selection

To identify the gene expression marker in the dataset GSE80655, the Genetic ALGOritms (GALGO) R package was used to classify the patients with MDD and HC, with the classification approach of ‘nearest centroid’ (Trevino and Falciani, 2006). GALGO is a genetic algorithm based on evolutionary-like principles. It combines genes, in this case, their expression, to find the best combination to classify any given classes (HC and MDD), with the highest possible accuracy. A chromosome in GALGO is defined as a combination of genes (biomarker size); a generation is the process of internal combinations of chromosomes, which is controlled by the user with a maximum number of solutions and a goal of

fitness (accuracy). The user can control the desired fitness, and the number of solutions if that fitness is not reached during the specified generation. If the accuracy (fitness) is not reached, and the max solutions target is reached, GALGO moves on to the next generation, and the process continues until it reaches the generations defined by the user, in our case 300.

An internal validation of two-thirds of the sample for training and one-third for testing was done to avoid overfitting. The accuracy of the gene expression marker was used to measure the statistical performance to classify between the two groups. The accuracy is the average between sensitivity and specificity, which is defined as  $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$ , where TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative.

The configuration of GALGO used chromosome size 10 (10-gene marker size), max solutions 300 and goal fitness 0.9. Also, to estimate the final error, the bootstrap method was used with 0.632 weight for test and 0.368 for train (Bradley and Tibshirani, 1993). The training error was a 10-fold cross-validation and the first-level splitting selected was 0.5 and 0.5 in only the 20 first splits. We used a forward selection model (FSM) algorithm only in dataset GSE80655, because this was potentially the most informative one because of its larger sample, greater number of brain structures included and use of RNASeq data. FSM is a method of choosing the expression of the genes based on their contribution in improving the classification accuracy. The process starts with the gene that has the highest accuracy, meaning a univariate classification between MDD and HC. Then, each next gene that significantly improves accuracy is added until there is no further improvement.

Finally, a heatmap of the gene expression signature was generated to visually display the difference in the gene expression between MDD and HC subjects, using *gplots* R Package, with the data *z*-scaled per row.

#### **Robustness and reproducibility capacity tests**

The robustness and reproducibility tests were implemented in brain and blood datasets. Reproducibility (also referred to as replicability) refers to obtaining the same results from the conduct of an independent study whose procedures are as closely matched to the original experiment as possible. Robustness refers to the stability of experimental conclusions to variations in either baseline assumptions or experimental procedures. It is somewhat related to the concept of generalizability (also known as transportability), which refers to the persistence of an effect in settings different from and outside of an experimental framework (Goodman *et al.*, 2016).

Due to the lack of a quantitative definition of reproducibility and robustness in the literature, we defined some parameters. Reproducibility was considered high, moderate or low, taking into account the SD calculated from the accuracy result in all datasets. Hence, high reproducibility is an accuracy within 1 SD, moderate within 2 SDs and low within 3 SDs or higher, in comparison with the accuracy of the normative dataset GSE80655 (which was 0.8).

We defined robustness as high when more than 90% of the independent datasets achieve reproducibility within 1 SD in comparison with the accuracy of GSE80655. For moderate and low robustness, we considered that more than 90% of the independent datasets achieved reproducibility within 2 and 3 SDs, respectively, in comparison with the accuracy of GSE80655.

#### **Correlation analysis between samples in brain and blood datasets**

To examine a relationship between expression in brain and blood tissue samples using the shared genes of each dataset of the identified marker, an analysis of correlation was calculated in R with the function *Pearson's*, creating a correlation matrix. The correlation matrix is presented as a heatmap with the function *heatmap.2* of *gplots* R package.

#### **Enrichment analysis of the identified marker**

Finally, to investigate the biological implications and significance of the identified gene expression marker, a Gene Ontology (GO) analysis was performed using GO terms on the STRING online database (<https://string-db.org>).

STRING determines protein–protein interaction by genomic context (neighborhood, fusion and co-occurrence), genomic experiments (co-expression and experiments), and prior consolidated knowledge on protein–protein associations (knowledge and text-mining).

In the case of association by neighborhood, two proteins are given an association score when their encoding genes are in close proximity to each other on the chromosome. For the fusion channel, all genomes are scanned for open reading frames that appear to be the result of gene fusion events. In co-occurrence, pairs of genes are identified whose occurrence patterns throughout evolution show similarities. For co-expression, pairs of genes that show consistent similarities between their expression profiles are assigned association scores. The experiments collect protein–protein interaction evidence from experiments and assays in the laboratory. Knowledge channel parses association evidence from curated pathway databases (i.e. Kyoto Encyclopedia of Genes and Genomes and Reactome), where it has been collected and consolidated manually by expert curators. Finally, the text-mining channel is based on PubMed abstracts and other sources. Pairs of proteins mentioned together in the same sentence, the same paragraph or merely the same publication are assigned a benchmarked association score (Szklarczyk *et al.*, 2021).

## **Results**

### **Identification of brain and blood datasets on Gene Expression Omnibus DataSets of NCBI**

For the current in-silico analysis, 21 datasets were identified on the GEO Datasets of NCBI (Table 1), 14 from brain and 7 from blood tissues. The clinical data can be reviewed in Supplementary Table 1, Supplemental digital content 1, <http://links.lww.com/PG/A260>.

The datasets GSE54562 and GSE54563 were merged and analyzed as a one dataset. Merging was also done for the datasets GSE54567 with GSE54568, and GSE54565 with GSE54572, resulting in three datasets instead of six. This was done because the datasets shared the same microarrays platform to measure gene expression, as well as the same brain structures (Table 1). This increased the sample size in the three final datasets and decreased the internal variance.

### **Analysis and filtering of variables**

The gene expression marker was identified in the dataset GSE80655, because it had more diverse brain areas, and the RNA levels were measured with RNASeq, which is more accurate than microarrays. From the 57 905 total probes of the dataset, 56 658 probes corresponded genes with protein product, 958 to pseudogenes, 1 to an RNA long non-coding and 288 with unknown biological implications. Twenty-eight genes were differentially expressed between MDD and HC.

### **GALGO R package for marker identification and feature selection**

After the filtering process was completed, the 28 genes were analyzed with GALGO R package. As a result, 23 genes were selected as the most important to differentiate



between patients with MDD and HC, with an accuracy of 0.8 (Table 1).

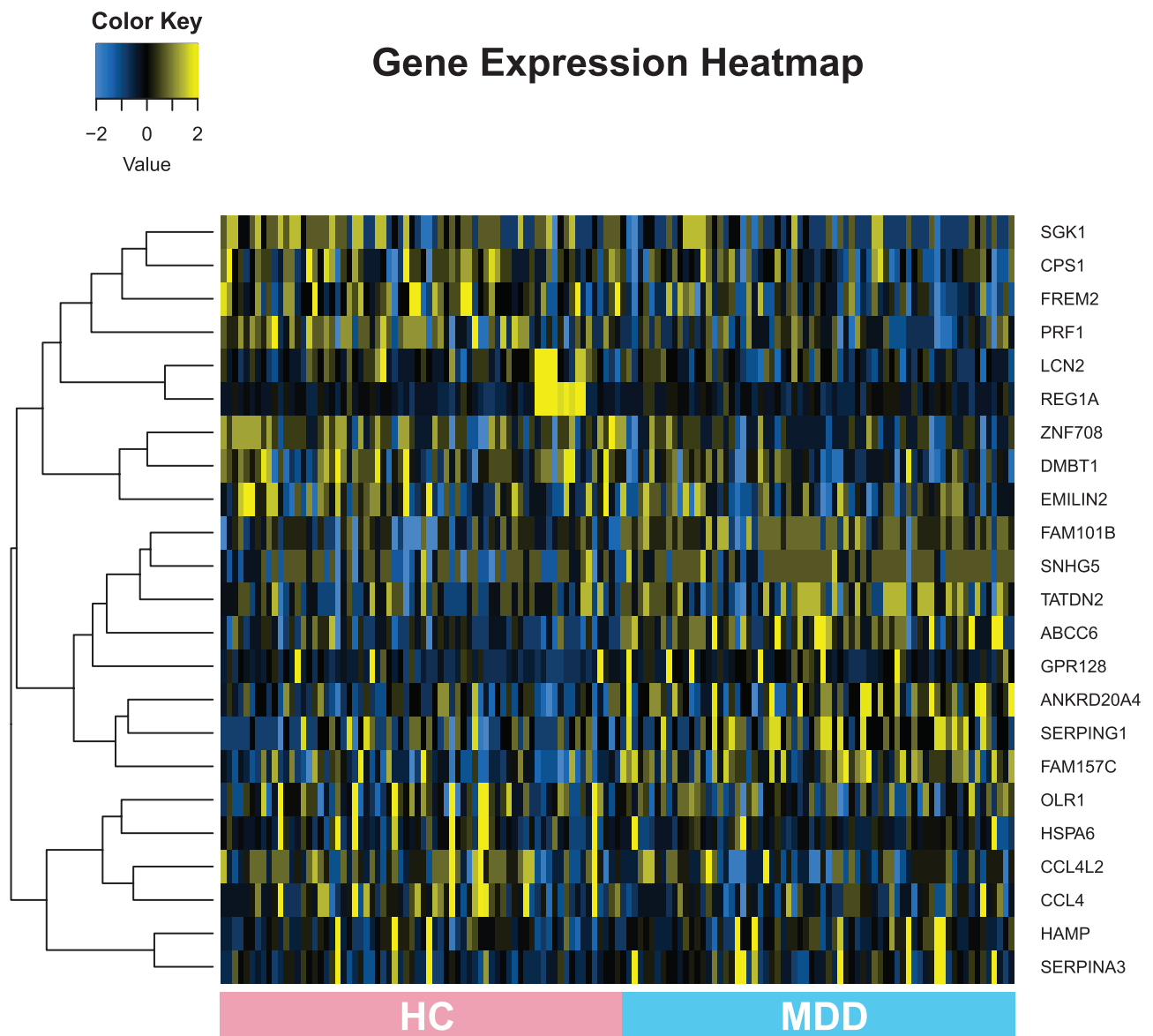
### Robustness and reproducibility capacity tests

The results of the accuracy of the tests are presented in Table 1 for each dataset. Figure 2 illustrates the heatmap of the gene expression differences across the samples in the MDD and HC classes.

The dataset GSE38206 had gene expression of subjects both during baseline depression and 8 weeks later after remission. In the present analysis, only the baseline

sample was included. GSE87610 included postmortem data in schizophrenia, bipolar disorder, MDD and HC. Gene expression was from pyramidal neurons in layers 3 and 5 of the dorsolateral prefrontal cortex (DLPFC). We averaged expression from both layers because they were highly correlated in both MDD and HC. GSE76826 examined gene expression in late-onset depression (LOD) at baseline, the same subjects in remission and HCs. We included only the baseline data of LOD as well as HCs. The dataset GSE19738 originally compared baseline and activation of the PBMCs to lipopolysaccharide

Fig. 2



Heatmap with a hierarchical clustering of the gene expression of the 23 gene marker in the dataset GSE80655. The classes are MDD (subjects with major depressive disorder) and HC (healthy control), and are grouped by diagnosis in the *x* axis. The genes, grouped hierarchically in the *y* axis, have a yellow color if the expression is high in that subject, whereas the genes with low expression have a blue color.

in subjects with and without depression. We used gene expression of PBMC before activation. Finally, in the dataset GSE67663, the subjects had comorbidity of depression and posttraumatic stress disorder.

The single SD of the accuracy for all datasets was 0.096; for 2 SDs it was 0.19, and for 3 SDs it was 0.29. Hence, datasets with an accuracy between 0.8 and 0.7 had high reproducibility; those with an accuracy of 0.61–0.69, moderate reproducibility; and those with less than 0.6 were considered to have low reproducibility. The only dataset with high reproducibility was GSE38206. The datasets GSE87610, GSE101521, GSE76826 and GSE39653 had moderate reproducibility, and the rest (12 datasets) had low reproducibility.

Based on our definition of robustness, our marker had a low robustness performance. Only 5 of the 17 datasets had a high or moderate reproducibility (30%). However, datasets GSE54564, GSE44593, GSE54570 and GSE98793 had accuracies between 0.58 and 0.59, very close to the 0.61 threshold for moderate reproducibility. Hence, the inclusion of these four datasets improved the overall robustness to 53% with moderate or high reproducibility.

#### Correlation analysis between samples in brain and blood datasets

Several complementary approaches were used to examine the correlation of gene expression in brain and blood tissues. First, we explored the impact of the platform used (microarray or RNASeq). The datasets GSE80655 (DLPFC measured with RNASeq), GSE101521 (DLPFC measured by RNASeq), and GSE87610 (DLPFC measured with microarrays) were selected because they shared more genes from the 23 gene marker. Results are depicted in Fig. 3 showing a high correlation between the genes [correlation mean = 0.7 (range = 0.16–1)].

Second, a correlation between brain samples (datasets GSE80655, GSE87610, GSE101521 and GSE53987), and selected PBMC samples (datasets GSE38206, GSE76826, GSE39653, GSE52790 and GSE98793) was done, including 14 shared genes in the original 23 gene marker. The results are depicted in Fig. 4 showing a higher linear correlation with Pearson coefficients between samples in different tissues; specifically, the brain datasets GSE87610 and GSE53987, with the blood datasets GSE39653 and GSE38206 with an average correlation of 0.38 (range = –0.3 to 0.8). The selection of these datasets was based on (a) higher accuracy in comparison with the other datasets (greater than or equal to 0.61; hence, at least moderate reproducibility); (b) higher number of genes present in the dataset corresponding to the marker (greater than or equal to 18 genes); or (c) higher number of samples in the datasets ( $n \geq 100$ ; this size facilitates the testing of correlations among samples independent of accuracy). More specifically, GSE80655, GSE87610, GSE38206, GSE76826 and GSE39653 were

selected due to their higher accuracy. (i.e. greater than or equal to 0.61). GSE101521 was selected because the 23 genes of the marker were present in the dataset and also measured the gene expression with RNASeq. GSE52790 was included based on the number of genes of the bio-marker present in the dataset (19/23). Finally, GSE53987 and GSE98793 were selected because of their particularly large samples, 105 and 182, respectively.

Finally, there was no statistical difference between brain and blood samples in the accuracy of the gene expression set we found using a *t* test ( $t = -0.36$ ,  $P = 0.77$ ).

#### Enrichment analysis of the identified marker

The results from the GO analysis are presented in Table 2.

#### Discussion

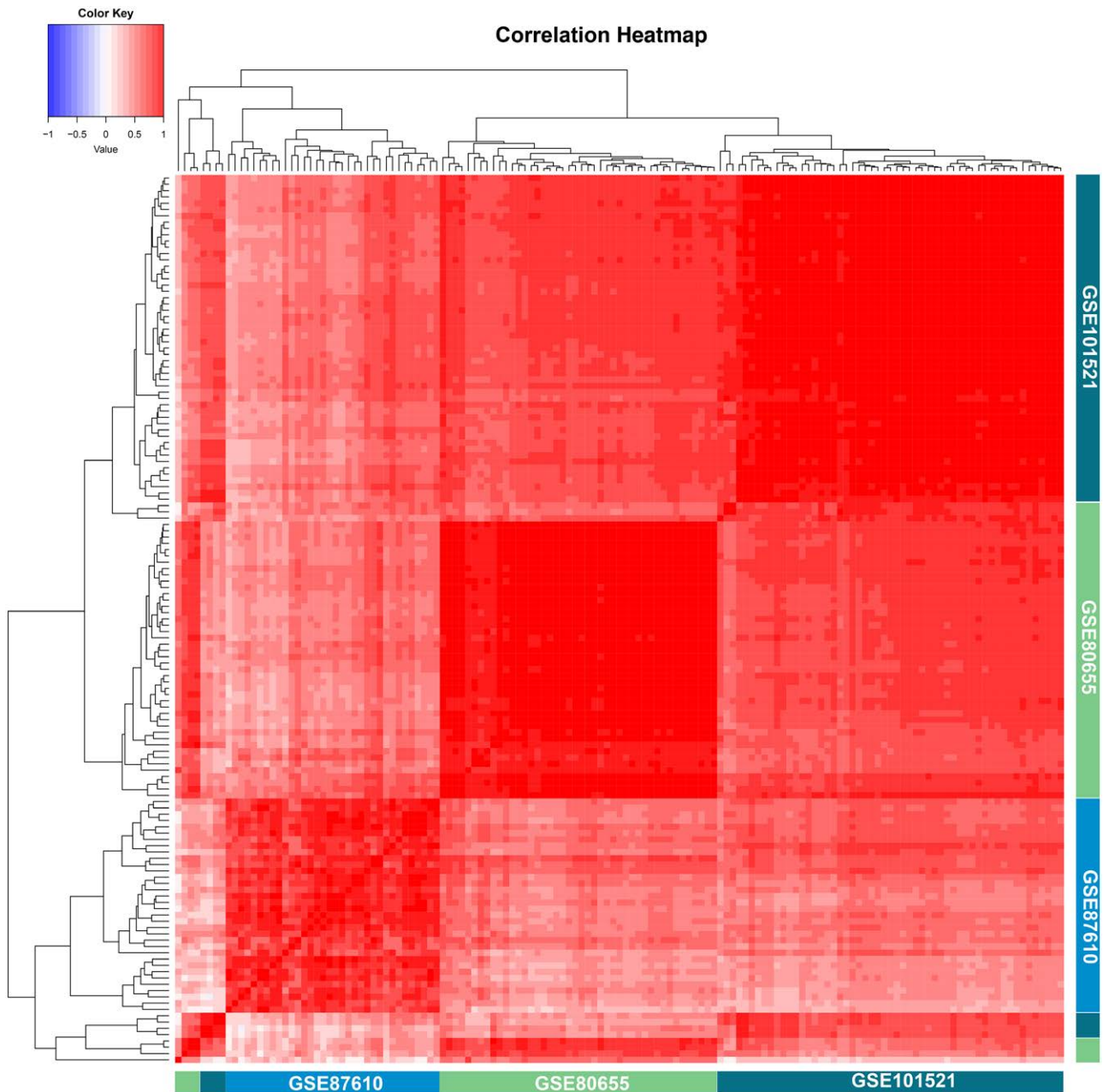
We examined all the available data up to October 2019 in the GEO Datasets <https://www.ncbi.nlm.nih.gov/geo/>, from subjects with MDD and HC and from brain and blood tissues. We identified a set of 23 protein-coding genes, capable of distinguishing MDD from HC with an average accuracy greater than 0.50, in 7 of 11 brain and in 6 of 7 blood datasets. Furthermore, the expression of 14 of the 23 genes was correlated in some brain and blood tissue samples. In the enrichment analysis, the identified genes were associated with immune response, a pathophysiological mechanism known to be involved in the development of depression. It is worth highlighting that the majority of the databases employed in this study were not originally used for the distinction of MDD from HCs.

Several studies have used gene expression sets to distinguish MDD from HC groups. Yi *et al.* (2012) examined PBMCs with microarrays in MDD ( $n = 8$ ), HC ( $n = 8$ ), and subsyndromal symptomatic depression ( $n = 8$ ), and reported a model of 48 genes with an accuracy of 100%, using a leave-one-out cross-validation (LOOCV) approach. However, LOOCV is known to be inconsistent with the linear model assumptions used. Hence, this study was limited by a very small sample and lack of validation in an independent dataset.

Wang *et al.* (2019) originally used dataset GSE98793 (included in our analyses) which they split into two subgroups of HC ( $n = 32$  each) and two of MDD ( $n = 64$  each). They identified a group of 20 hub genes (genes with high protein–protein interactions with other genes) in PBMCs capable of classifying MDD from HC, with an accuracy (from receiving operating curve analyses) between 0.833 and 0.876 with validation in the second subgroups.

Miyata *et al.* (2016) studied late-onset MDD (LOD; in subjects aged greater than or equal to 50 years;  $n = 18$ ) and HC ( $n = 12$ ). They first examined a 3066 probes microarray in PBMCs from a mice model of depression.

Fig. 3



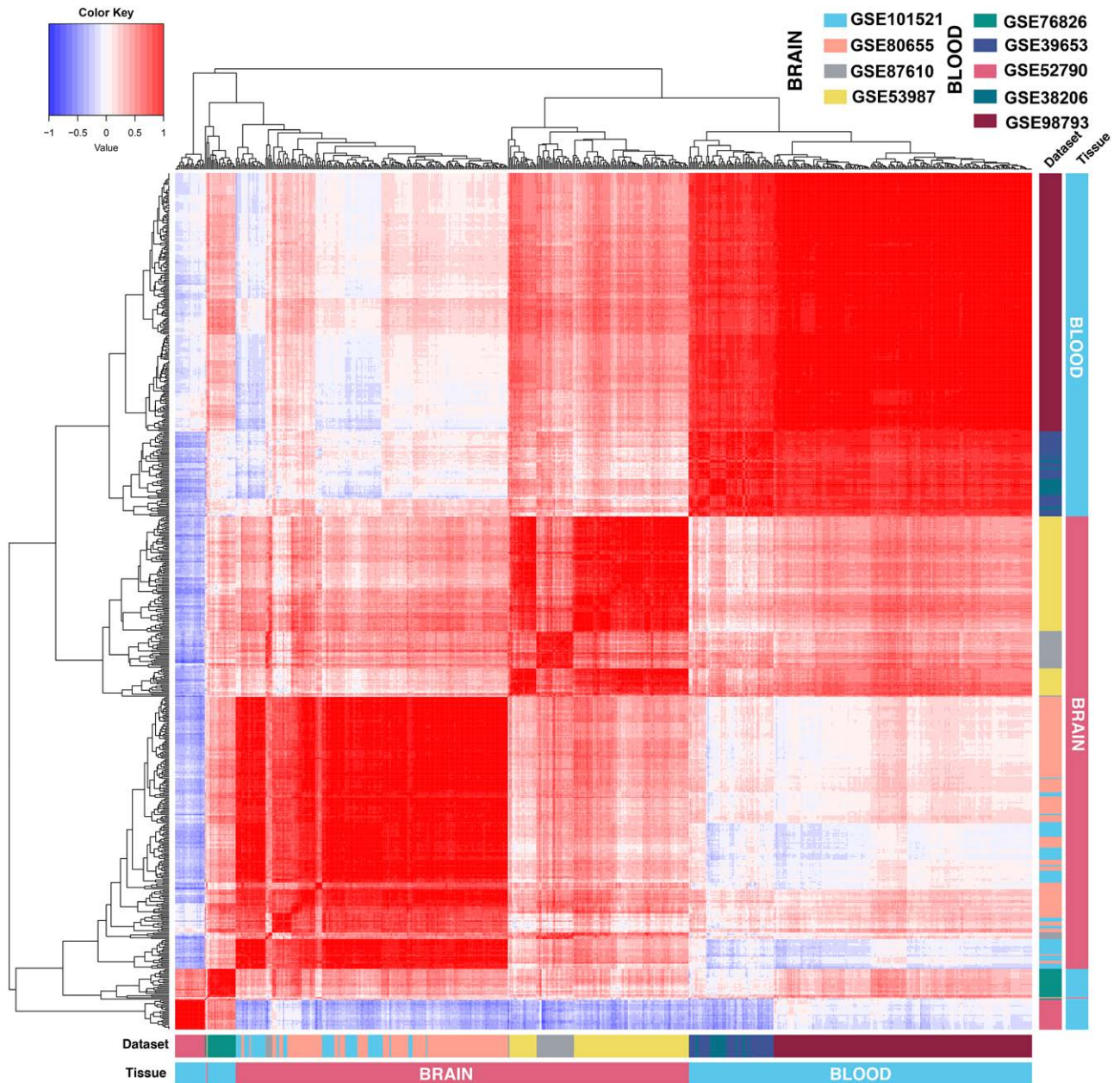
Heatmap of the correlation between the subjects MDD and HC in the datasets GSE80655, GSE101521 and GSE87610 of only the DLPFC of the mentioned datasets. This correlation was calculated using the gene expression of 20 of the 23 genes of original the marker, due to these are the shared genes in these three datasets. The purpose of correlating the data of RNASeq (GSE80655 and GSE101521) and microarrays (GSE87610) in the same brain area (DLPFC) was to test if there is a correlation dependent of the used platform to measure the gene expression. The correlations are hierarchically grouped in  $x$  and  $y$  axes. A positive linear correlation is represented as red color whereas a negative linear correlation is blue; the color white is used to represent the lack of linear correlation. DLPFC, dorsolateral prefrontal cortex; HC, healthy control; MDD, major depressive disorder.

Fourteen genes were differentially expressed and concordant between the animal model and LOD subjects. Finally, the ability to classify LOD and HC was tested using reverse transcription polymerase chain reaction

(RT-PCR) in independent LOD subjects ( $n = 23$ ). The *CIDE*C gene (cell death-inducing DFFA-like effector C) reached the highest accuracy (0.9185). However, their sample was small.



Fig. 4



Heatmap of the correlation using 14 of the 23 genes from the original marker, between the subjects MDD and HC, between the datasets of brain tissue GSE80655, GSE87610, GSE101521 and GSE53987 and the PBMC datasets GSE38206, GSE76826, GSE39653, GSE52790 and GSE98793. The correlations are hierarchically grouped in  $x$  and  $y$  axes. A positive linear correlation is represented as red color whereas a negative linear correlation is blue; the color white is used to represent the lack of linear correlation. HC, healthy control; MDD, major depressive disorder; PBMC, peripheral blood mononuclear cell.

Spijker *et al.* (2010) studied whole blood with microarrays from a primary cohort of MDD ( $n = 21$ ) and HC ( $n = 21$ ). They reported a set of seven genes differentially expressed. These findings were validated in an independent cohort of MDD ( $n = 13$ ) and HC ( $n = 14$ ) using RT-PCR. The reported specificity was 71.4% and sensitivity was 76.9%; accuracy was not reported. As was

already mentioned, small studies tend to inflate effects (Button *et al.*, 2013).

Bhak *et al.* (2019) used methylome acquired with methyl-seq data and transcriptome obtained with whole-transcriptome sequencing from PBMCs of subjects with a history of suicidal attempt (SA,  $n = 56$ ), MDD ( $n = 39$ )



**Table 2 Gene Ontology analysis of the gene expression marker**

Term description	Observed gene count	Background gene count	False discovery rate	Matching proteins in your network (labels) gene ID
Response to stress	13	3267	0.0013	CCL4, CCL4L1, CPS1, DMBT1, GIG25, HAMP, HSPA6, LCN2, OLR1, PRF1, SERPING1, SGK1, TATDN2
Defense response	9	1234	0.0013	CCL4, CCL4L1, DMBT1, GIG25, HAMP, LCN2, OLR1, PRF1, SERPING1
Immune response	10	1560	0.0013	CCL4, CCL4L1, DMBT1, GIG25, HAMP, HSPA6, LCN2, OLR1, PRF1, SERPING1
Immune effector process	7	927	0.0072	DMBT1, GIG25, HSPA6, LCN2, OLR1, PRF1, SERPING1

and HC ( $n = 87$ ). They first reported differentially methylated sites (DMSs) for each clinical group versus HC. The MDD had 80 DMS and the SA group had 95 compared to HCs. Gene expression models failed to differentiate the clinical groups from HCs. However, gene expression combined with DMS achieved accuracies of 87.3% for MDD versus HC and 86.7% for SA versus HC. This underlines the potential advantage of linking gene expression with epigenetic markers.

Watanabe *et al.* (2015) through PCR array plates in a primary dataset assessed 40 candidate genes in PBMCs from subjects with MDD ( $n = 25$ ) and HC ( $n = 25$ ). A group of seven genes achieved a sensitivity of 80% and specificity of 92% in the primary dataset. In an independent dataset test, including subjects with MDD ( $n = 20$ ) and HC ( $n = 18$ ), the specificity and sensitivity were 85 and 89%, respectively. Accuracy was not reported.

This literature has not succeeded in the identification of a reliable gene expression biomarker for MDD in the blood (Spijker *et al.*, 2010; Yi *et al.*, 2012; Watanabe *et al.*, 2015; Miyata *et al.*, 2016; Bhak *et al.*, 2019; Wang *et al.*, 2019). The studies have some limitations such as using LOOCV as an internal test (Yi *et al.*, 2012; Bhak *et al.*, 2019) which is inconsistent when used in linear models (Shao, 1993); subsplitting the data instead of using an independent dataset for validation (Wang *et al.*, 2019); and use of small datasets (Spijker *et al.*, 2010; Watanabe *et al.*, 2015; Miyata *et al.*, 2016; Bhak *et al.*, 2019) which are biased in favor of inflated effects (Button *et al.*, 2013). We found no reports using brain tissue for the diagnostic classification of MDD.

Our accuracy findings are similar to the studies described above (0.8 in the primary brain dataset, with a range between 0.46 and 0.8 in the other brain datasets and between 0.45 and 0.78 for the blood datasets). However, our approach differed from others in that the expression set was restricted to protein-coding genes, first identified in brain and then replicated in both brain and blood. This is important because expression in brain supports the neurobiological validity of the marker, whereas expression in blood supports its potential to be eventually developed as a diagnostic aid. Conceptually, the fact that the GO analyses link the identified marker to immune, stress and defense responses is consistent with the broad literature supporting the involvement of the immune

system in the pathophysiology of depression (Stein *et al.*, 1991; Maes *et al.*, 1995; Raison *et al.*, 2006, 2009; Quan and Banks, 2007; Dantzer *et al.*, 2008; Dowlati *et al.*, 2010; Leonard, 2010; Hughes *et al.*, 2016; Miller and Raison, 2016). Pragmatically, because ours is a compact marker, the implementation of RT-PCR in blood seems feasible. Hence, we believe a further study of this marker in a larger dataset with PCR is warranted.

Some technical aspects negatively impacted the performance of the identified marker. First, in the brain tissue datasets, different regions were included. In brain, there tends to be high variability of gene expression in different areas (Naumova *et al.*, 2013). However, the identified marker was capable of distinguishing MDD from HC in most of the brain datasets, independently of the included brain areas (Table 1). Moreover, postmortem interval affects the RNA levels and integrity (Franz *et al.*, 2005) which could be accountable for the lower accuracy in GSE10152, even though, this dataset used RNASeq samples from the DLPFC area, like in GSE80655. Still, the gene set was able to discriminate the groups with an accuracy greater than 0.6. Additionally, inconsistency with the previous literature may, in part, be due to greater tolerance to degradation in the gene expression we identified in GSE80655, compared to some of the genes previously reported. Also, the inherent variability of microarray experiments is large (Jaksik *et al.*, 2015). As initially detected in GSE80655, we expected greater gene expression differences between MDD and HC in the replication datasets. However, although the microarray probes measured the same genes, the variability of the probes was not consistent across the datasets. We mitigated this with the selection of the probes that reflect the original assumption, with an expression difference between MDD and HC. Furthermore, the microarray experiments used different number of probes (up to 48 for GSE87610 and as low as 12 for GSE67663). This may account for the higher accuracy in GSE98793 compared to GSE67663, although the sample sizes were almost the same. In general, datasets with more probes, such as GSE38206 and GSE76826, had a higher accuracy, but this could be artifactual. Of note, variability in sample sizes was significant (e.g. 192 in GSE98793 and 18 in GSE38206). Smaller datasets achieved higher accuracy perhaps due to higher internal variability, biasing toward inflated effects (Button *et al.*, 2013). Finally, the association of gene expression between

brain and blood samples was implemented in tissues originating from different subjects. Ideally, samples would come from the same subject collected at the same time (Sullivan *et al.*, 2006; Leday *et al.*, 2018). This would reduce variation in clinical characteristics that could impact expression (Ciobanu *et al.*, 2016; Yang *et al.*, 2017).

The current analysis showed that (a) there is an association between the gene expression of the identified marker from subjects with MDD and HC in the blood and brain tissues; (b) a 23 gene expression marker was able to distinguish subjects with MDD from HC with an accuracy higher than 50% in most of the databases investigated and (c) this marker includes genes related to stress and immune response, consistent with a current model of depression. Future studies in larger samples will be necessary, including brain and blood samples from the same subjects to further develop a marker with minimal sample bias, which is reproducible and universal. Though reliability between clinicians diagnosing MDD according to DSM-5 criteria was poor (Regier *et al.*, 2013), it improved with the Structured Clinical Interview for DSM-5 – Clinician Version (SCID-5-CV; kappa greater than 0.70 for MDD) (Osório *et al.*, 2019). Hence, a future study would examine the accuracy of our proposed genetic biomarker compared to the SCID-5-CV as a gold standard.

## Acknowledgements

### Conflicts of interest

Juan Bustillo receives royalties from UpToDate. For the remaining author, there are no conflicts of interest.

## References

- Arion D, Huo Z, Enwright JF, Corradi JP, Tseng G, Lewis DA (2017). Transcriptome alterations in prefrontal pyramidal cells distinguish schizophrenia from bipolar and major depressive disorders. *Biol Psychiatry* **82**:594–600.
- Bahn S, Augood SJ, Ryan M, Standaert DG, Starkey M, Emson PC (2001). Gene expression profiling in the post-mortem human brain: no cause for dismay. *J Chem Neuroanat* **22**:79–94.
- Behera P, Gupta SK, Nongkynrih B, Kant S, Mishra AK, Sharan P (2017). Screening instruments for assessment of depression. *Indian J Med Spec* **8**:31–37.
- Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* **57**: 289–300.
- Bhak Y, Jeong HO, Cho YS, Jeon S, Cho J, Gim JA, *et al.* (2019). Depression and suicide risk prediction models using blood-derived multi-omics data. *Transl Psychiatry* **9**:262.
- Bradley E, Tibshirani RJ (2000). *An introduction to the bootstrap*. Chapman&Hall/CRC.
- Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* **14**:365–376.
- Canuto A, Gkinis G, DiGiorgio S, Arpone F, Herrmann FR, Weber K (2016). Agreement between physicians and liaison psychiatrists on depression in old age patients of a general hospital: influence of symptom severity, age and personality. *Aging Ment Health* **20**:1092–1098.
- Chang LC, Jamain S, Lin CW, Rutjescu D, Tseng GC, Sibille E (2014). A conserved BDNF, glutamate- and GABA-enriched gene module related to human depression identified by coexpression meta-analysis and DNA variant genome-wide association studies. *PLoS One* **9**:e90980.
- Ciobanu LG, Sachdev PS, Trollor JN, Reppermund S, Thalamuthu A, Mather KA, *et al.* (2016). Differential gene expression in brain and peripheral tissues in depression across the life span: a review of replicated findings. *Neurosci Biobehav Rev* **71**:281–293.
- Dantzer R, O'Connor JC, Freund GG, Johnson RW, Kelley KW (2008). From inflammation to sickness and depression: when the immune system subjugates the brain. *Nat Rev Neurosci* **9**:46–56.
- Deep-Soboslay A, Benes FM, Haroutunian V, Ellis JK, Kleinman JE, Hyde TM (2011). Psychiatric brain banking: three perspectives on current trends and future directions. *Biol Psychiatry* **69**:104–112.
- Dowlati Y, Herrmann N, Swardfager W, Liu H, Sham L, Reim EK, Lanctôt KL (2010). A meta-analysis of cytokines in major depression. *Biol Psychiatry* **67**:446–457.
- Enns MW, Larsen DK, Cox BJ (2000). Discrepancies between self and observer ratings of depression. The relationship to demographic, clinical and personality variables. *J Affect Disord* **60**:33–41.
- Erdő F, Denes L, de Lange E (2017). Age-associated physiological and pathological changes at the blood-brain barrier: a review. *J Cereb Blood Flow Metab* **37**:4–24.
- Franz H, Ullmann C, Becker A, Ryan M, Bahn S, Arendt T, *et al.* (2005). Systematic analysis of gene expression in human brains before and after death. *Genome Biol* **6**:R112.
- Friedrich MJ (2017). Depression is the leading cause of disability around the world. *JAMA* **317**:1517.
- Gadad BS, Jha MK, Czystz A, Furman JL, Mayes TL, Emslie MP, Trivedi MH (2018). Peripheral biomarkers of major depression and antidepressant treatment response: current knowledge and future outlooks. *J Affect Disord* **233**:3–14.
- Gao S, Calhoun VD, Sui J (2018). Machine learning in major depression: from classification to treatment outcome prediction. *CNS Neurosci Ther* **24**:1037–1052.
- Goodman SN, Fanelli D, Ioannidis JP (2016). What does research reproducibility mean? *Sci Transl Med* **8**:341ps12.
- Hacimusalar Y, Eşel E (2018). Suggested biomarkers for major depressive disorder. *Noro Psikiyatir Ars* **55**:280–290.
- Harrison PJ (2011). Using our brains: the findings, flaws, and future of postmortem studies of psychiatric disorders. *Biol Psychiatry* **69**:102–103.
- Hughes MM, Connor TJ, Harkin A (2016). Stress-related immune markers in depression: implications for treatment. *Int J Neuropsychopharmacol* **19**:pyw001.
- Jaksik R, Iwanaszko M, Rzeszowska-Wolny J, Kimmel M (2015). Microarray experiments and factors which affect their reliability. *Biol Direct* **10**:46.
- Leday GGR, Vértés PE, Richardson S, Greene JR, Regan T, Khan S, *et al.*; MRC Immunopsychiatry Consortium (2018). Replicable and coupled changes in innate and adaptive immune gene expression in two case-control studies of blood microarrays in major depressive disorder. *Biol Psychiatry* **83**:70–80.
- Leonard BE (2010). The concept of depression as a dysfunction of the immune system. *Curr Immunol Rev* **6**:205–212.
- Li JZ, Vawter MP, Walsh DM, Tomita H, Evans SJ, Choudary PV, *et al.* (2004). Systematic changes in gene expression in postmortem human brains associated with tissue pH and terminal medical conditions. *Hum Mol Genet* **13**:609–616.
- Liu Z, Li X, Sun N, Xu Y, Meng Y, Yang C, *et al.* (2014). Microarray profiling and co-expression network analysis of circulating lncRNAs and mRNAs associated with major depressive disorder. *PLoS One* **9**:e93388.
- Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550.
- Maes M, Meltzer HY, Bosmans E, Bergmans R, Vandoolaeghe E, Ranjan R, Desnyder R (1995). Increased plasma concentrations of interleukin-6, soluble interleukin-6, soluble interleukin-2 and transferrin receptor in major depression. *J Affect Disord* **34**:301–309.
- McCullumsmith RE, Meador-Woodruff JH (2011). Novel approaches to the study of postmortem brain in psychiatric illness: old limitations and new challenges. *Biol Psychiatry* **69**:127–133.
- Miller AH, Raison CL (2016). The role of inflammation in depression: from evolutionary imperative to modern treatment target. *Nat Rev Immunol* **16**:22–34.
- Miyata S, Kurachi M, Okano Y, Sakurai N, Kobayashi A, Harada K, *et al.* (2016). Blood transcriptomic markers in patients with late-onset major depressive disorder. *PLoS One* **11**:e0150262.
- Naumova OY, Lee M, Rychkov SY, Vlasova NV, Grigorenko EL (2013). Gene expression in the human brain: the current state of the study of specificity and spatiotemporal dynamics. *Child Dev* **84**:76–88.
- NIMH (2020). *Major Depression*. <https://www.nimh.nih.gov/health/statistics/major-depression.shtml> [Accessed 30 April 2020]
- Osório FL, Loureiro SR, Hallak JEC, Machado-de-Sousa JP, Ushirohira JM, Baes CVW, *et al.* (2019). Clinical validity and intrarater and test-retest reliability of the Structured Clinical Interview for DSM-5 - Clinician Version (SCID-5-CV). *Psychiatry Clin Neurosci* **73**:754–760.
- Pantazatos SP, Huang YY, Rosoklija GB, Dwork AJ, Arango V, Mann JJ (2017). Whole-transcriptome brain expression and exon-usage profiling in major

- depression and suicide: evidence for altered glial, endothelial and ATPase activity. *Mol Psychiatry* **22**:760–773.
- Quan N, Banks WA (2007). Brain-immune communication pathways. *Brain Behav Immun* **21**:727–735.
- Raison CL, Borisov AS, Majer M, Drake DF, Pagnoni G, Woolwine BJ, *et al.* (2009). Activation of central nervous system inflammatory pathways by interferon-alpha: relationship to monoamines and depression. *Biol Psychiatry* **65**:296–303.
- Raison CL, Capuron L, Miller AH (2006). Cytokines sing the blues: inflammation and the pathogenesis of depression. *Trends Immunol* **27**:24–31.
- Ramaker RC, Bowling KM, Lasseigne BN, Hagenauer MH, Hardigan AA, Davis NS, *et al.* (2017). Post-mortem molecular profiling of three psychiatric disorders. *Genome Med* **9**:72.
- Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, Kupfer DJ (2013). DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses. *Am J Psychiatry* **170**:59–70.
- Savitz J, Frank MB, Victor T, Bebak M, Marino JH, Bellgowan PS, *et al.* (2013). Inflammation and neurological disease-related genes are differentially expressed in depressed patients with mood disorders and correlate with morphometric and functional imaging abnormalities. *Brain Behav Immun* **31**:161–171.
- Shao J (1993). Linear model selection by cross-validation. *J Am Stat Assoc* **88**:486–494.
- Spijker S, Van Zanten JS, De Jong S, Penninx BW, van Dyck R, Zitman FG, *et al.* (2010). Stimulated gene expression profiles as a blood marker of major depressive disorder. *Biol Psychiatry* **68**:179–186.
- Stein M, Miller AH, Trestman RL (1991). Depression, the immune system, and health and illness. Findings in search of meaning. *Arch Gen Psychiatry* **48**:171–177.
- Strawbridge R, Young AH, Cleare AJ (2017). Biomarkers for depression: recent insights, current challenges and future prospects. *Neuropsychiatr Dis Treat* **13**:1245–1262.
- Sullivan PF, Daly MJ, Ripke S, Lewis CM, Lin DY, *et al.* (2013). A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* **18**:497–511.
- Sullivan PF, Fan C, Perou CM (2006). Evaluating the comparability of gene expression in blood and brain. *Am J Med Genet B Neuropsychiatr Genet* **141B**:261–268.
- Sullivan PF, Neale MC, Kendler KS (2000). Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* **157**:1552–1562.
- Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, *et al.* (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**:D605–D612.
- Tolentino JC, Schmidt SL (2018). DSM-5 criteria and depression severity: implications for clinical practice. *Front Psychiatry* **9**:450.
- Tomita H, Vawter MP, Walsh DM, Evans SJ, Choudary PV, Li J, *et al.* (2004). Effect of agonal and postmortem factors on gene expression profile: quality control in microarray analyses of postmortem human brain. *Biol Psychiatry* **55**:346–352.
- Trevino V, Falciani F (2006). GALGO: an R package for multivariate variable selection using genetic algorithms. *Bioinformatics* **22**:1154–1156.
- Wang H, Zhang M, Xie Q, Yu J, Qi Y, Yue Q (2019). Identification of diagnostic markers for major depressive disorder by cross-validation of data from whole blood samples. *PeerJ* **7**:e7171.
- Watanabe SY, Iga J, Ishii K, Numata S, Shimodera S, Fujita H, Ohmori T (2015). Biological tests for major depressive disorder that involve leukocyte gene expression assays. *J Psychiatr Res* **66–67**:1–6.
- Wingo AP, Almlí LM, Stevens JS, Stevens JJ, Klengel T, Uddin M, *et al.* (2015). DICER1 and microRNA regulation in post-traumatic stress disorder with comorbid depression. *Nat Commun* **6**:10106.
- Yang C, Hu G, Li Z, Wang Q, Wang X, Yuan C, *et al.* (2017). Differential gene expression in patients with subsyndromal symptomatic depression and major depressive disorder. *PLoS One* **12**:e0172692.
- Yi Z, Li Z, Yu S, Yuan C, Hong W, Wang Z, *et al.* (2012). Blood-based gene expression profiles models for classification of subsyndromal symptomatic depression and major depressive disorder. *PLoS One* **7**:e31283.