

University of Groningen

Major depressive disorder is associated with changes in a cluster of serum and urine biomarkers

van Buel, Erin M; Meddens, Marcus J M; Arnoldussen, Eduard A; van den Heuvel, Edwin R; Bohlmeijer, Willem C; den Boer, Johan A; Muller Kobold, Anna; Boonman-de Winter, Leandra J M; van Rumpft, Dirk; Timmers, Lambertus F J

Published in:
Journal of Psychosomatic Research

DOI:
[10.1016/j.jpsychores.2019.109796](https://doi.org/10.1016/j.jpsychores.2019.109796)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

van Buel, E. M., Meddens, M. J. M., Arnoldussen, E. A., van den Heuvel, E. R., Bohlmeijer, W. C., den Boer, J. A., Muller Kobold, A., Boonman-de Winter, L. J. M., van Rumpft, D., Timmers, L. F. J., Veerman, M. F. A., Kamphuis, J. S., Gladkevich, A. V., Schoevers, R. A., Luiten, P. G. M., Eisel, U. L. M., Bosker, F. J., & Klein, H. C. (2019). Major depressive disorder is associated with changes in a cluster of serum and urine biomarkers. *Journal of Psychosomatic Research*, *125*, [109796].
<https://doi.org/10.1016/j.jpsychores.2019.109796>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



ELSEVIER

Contents lists available at ScienceDirect

Journal of Psychosomatic Research

journal homepage: www.elsevier.com/locate/jpsychores

Major depressive disorder is associated with changes in a cluster of serum and urine biomarkers

Erin M. van Buel^{a,c,1}, Marcus J.M. Meddens^{b,1}, Eduard A. Arnoldussen^b,
 Edwin R. van den Heuvel^e, Willem C. Bohlmeijer^f, Johan A. den Boer^{g,h}, Anna Muller Kobold^c,
 Leandra J.M. Boonman-de Winterⁱ, Dirk van Rumpft^j, Lambertus F.J. Timmers^k,
 Mattheus F.A. Veerman^l, Johannes S. Kamphuis^m, Anatoliy V. Gladkevich^c,
 Robert A. Schoevers^{c,d,*}, Paul G.M. Luiten^a, Ulrich L.M. Eisel^{a,c}, Fokko J. Bosker^{c,2},
 Hans C. Klein^{c,g,2}

^a Groningen Institute for Evolutionary Life Sciences (GeLIFES), University of Groningen, The Netherlands

^b Brainlabs BV, Deventer, The Netherlands

^c University of Groningen, University Medical Center Groningen, Department of Psychiatry, The Netherlands

^d Research School of Behavioral and Cognitive Neurosciences, University of Groningen, Groningen, The Netherlands

^e Department of Mathematics & Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands

^f Office for Neuropsychiatry, Psychotherapy and Expertise, Brummen, The Netherlands

^g Department of Nuclear Medicine and Molecular Imaging, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

^h PRA Health Sciences, Groningen, The Netherlands

ⁱ Amphia Academy, Amphia Hospital, Breda, The Netherlands

^j SHO Center for Medical Diagnostics, Velp, The Netherlands

^k General Practitioner's Office BERGH, 's Heerenberg, The Netherlands

^l STAR-MDC Laboratories, Rotterdam, The Netherlands

^m Department of Clinical Chemistry and Hematology, Gelre Hospital, Apeldoorn, The Netherlands

ARTICLE INFO

Keywords:

Major depressive disorder

Biomarkers

Biomarker panel

Urine

Permutation analysis

ELISA

ABSTRACT

Major Depressive Disorder (MDD) is a heterogeneous disorder with a considerable symptomatic overlap with other psychiatric and somatic disorders. This study aims at providing evidence for association of a set of serum and urine biomarkers with MDD. We analyzed urine and serum samples of 40 MDD patients and 47 age- and sex-matched controls using 40 potential MDD biomarkers (21 serum biomarkers and 19 urine biomarkers). All participants were of Caucasian origin. We developed an algorithm to combine the heterogeneity at biomarker level. This method enabled the identification of correlating biomarkers based on differences in variation and distribution between groups, combined the outcome of the selected biomarkers, and calculated depression probability scores (the "bio depression score"). Phenotype permutation analysis showed a significant discrimination between MDD and euthymic (control) subjects for biomarkers in urine ($P < .001$), in serum ($P = .02$) and in the combined serum plus urine result ($P < .001$). Based on this algorithm, a combination of 8 urine biomarkers and 9 serum biomarkers were identified to correlate with MDD, enabling an area under the curve (AUC) of 0.955 in a Receiver Operating Characteristic (ROC) analysis. Selection of either urine biomarkers or serum biomarkers resulted in AUC values of 0.907 and 0.853, respectively. Internal cross-validation (5-fold) confirmed the association of this set of biomarkers with MDD.

1. Introduction

Major depressive disorder (MDD), with a lifetime prevalence of

around 15%, is a major cause of disability in the western world [1,2]. Due to the heterogeneous nature of MDD and its symptomatic overlap with other psychiatric and somatic disorders diagnosis may be

* Corresponding author at: University Medical Center Groningen, Academic Centre of Psychiatry, Hanzplein 1, 9713 GZ Groningen, The Netherlands.

E-mail address: r.a.schoevers@umcg.nl (R.A. Schoevers).

¹ The first two authors contributed equally to this manuscript.

² The last two authors contributed equally to this manuscript.

complicated [3]. The heterogeneity of the disorder most probably also finds its roots in heterogeneous biological processes as described by Jentsch et al. [3]. Six hypotheses were formulated, being those that describe pathophysiological processes of the HPA-axis, inflammation and immune system alterations, monoamine dysfunction, oxidative stress/endothelial dysfunction, altered neurogenesis/neuroplasticity, and alterations of magnesium metabolism/mineral homeostasis. Within these hypotheses typical biomarkers linked to the pathophysiological processes were identified. In a novel method to discriminate between patients and controls we eventually selected a subset of biomarkers covering all important hypotheses of MDD pathophysiology [3–7]. In supplementary table S1 the scientific support for these hypotheses including a list of relevant biomarkers are summarized.

A biomarker is defined as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” [8]. Thus, a biomarker test for MDD may aid the clinician in making a correct diagnosis or predicting treatment response. Several biomarkers have been suggested for MDD, including cytokines (e.g. TNF α , IL-1 β), neurotrophic factors (e.g. BDNF, VEGF), and hormones (e.g. cortisol). However, none of these biomarkers fulfill the sensitivity and specificity criteria when used separately [3]. This may be in part due to the complicated underlying pathophysiology of MDD. An increasing body of evidence indicates that the underlying neurobiology of MDD likely involves a complex interplay of genetic factors, dysregulation of the hypothalamic-pituitary-adrenal (HPA) axis and other endocrine parameters, dysfunctions in the immune system and monoaminergic systems. Accordingly, single genetic, endocrinological, neurotransmitter-related or hormonal abnormalities are unlikely to discriminate patients with severe mood disorders from healthy people or patients with other psychiatric disorders. Combining a number of biomarkers reflecting the divergent dysfunctions in MDD, as designed in this study, might be a more fruitful approach [3].

A major problem in the identification of biomarkers associated with disease is the fact that biomarker values are not normally distributed, and distributions may be different in patients and healthy controls. When the distribution in patients and healthy controls differs in aspects other than the mean or median, difficulties arise for parametric and non-parametric testing. Examples in which regular parametric or non-parametric testing fails include a ceiling effect in one of the groups or differences in variance between groups not accompanied by differences in average. Variance information gets lost when not considering that each biomarker obeys different variance rules between cases and controls.

In this study we have used an analytic method described and submitted by Paus et al. [9] that addresses these pitfalls. This method is called Quantile Based Prediction (QBP) and is used to select those biomarkers that “behave” differently between cases and controls while not necessarily displaying a difference in average between both groups. It distinguishes the distributional tail behavior between cases and healthy controls. Subsequently, the best performing biomarkers were selected, and its validity tested. This method combines the outcome of the selected biomarkers and calculates a depression probability score discriminating depressed from euthymic subjects. The method provides for each subject a biometric depression score, tentatively named BDS (Bio-Depression-Score). Aim of the study is to provide evidence for a cluster of serum and/or urine biomarkers to be associated with MDD, reflecting the complex interaction of divergent pathophysiological mechanisms that underlies depression.

2. Methods

2.1. Patients

MDD patients were recruited in collaboration with general practitioners, psychiatric care organizations and through advertisements in

local and national newspapers. For definite inclusion a MINI diagnostic interview was performed to validate the DSM-IV diagnosis of MDD and concurrently rule out organic causes. Further inclusion criteria included: age 18–65, a 17-item Hamilton Depression Rating Scale (HAM-D) score equal or higher than 11 and informed consent. Exclusion criteria included pregnancy, presence of a confounding primary psychiatric disorder in the MINI interview among which alcohol or substance use disorders, inflammatory or systemic diseases, metabolic disorders or other disorders that might cause MDD. A total of 102 volunteers were included (51 MDD, 51 Controls) in which all biomarkers were tested. Non-Caucasian ethnicity (11 MDD, 4 Controls) was excluded, resulting in 40 MDD cases and 47 healthy controls of Caucasian ethnicity. MDD cases consisted of 19 with and 21 without anti-depressant medication. Healthy controls (HC) were recruited via general practitioners and advertisements in local and national newspapers. For definite inclusion the MINI diagnostic interview was performed. Axis II diagnosis was not investigated in healthy volunteers nor MD patients, because recruitment was primary done in general care/general population settings and not expected to confound our results. HAM-D was used to assess symptoms of depression and the Mini International Neuropsychiatric Interview (MINI) was conducted for objective validation of diagnosis. A non-clinician researcher trained by an experienced clinician in the use of these questionnaires executed all questionnaires and a second experienced psychiatrist checked the correct diagnosis.

Participants were asked to deliver 50 ml of blood through venipuncture as well as 50 ml of first morning urine. Urine and blood were collected the same day. Per participant a single blood sample was collected in the course of the day and documented in 66 of the 87 participants: from 8 am till 6 pm, mean 11:50 am and latest 5:40 pm. Blood was collected in serum separation tubes, allowed to clot and centrifuged at 3000 xg for 10 min. Serum supernatant was divided into aliquots and stored at -80°C . Urine samples were centrifuged for 10 min at 1000 xg to precipitate any particles and cells; the supernatant was collected, divided into aliquots and stored at -80°C .

2.2. ELISAs

ELISA kits were obtained from the following vendors: R&D systems Europe Ltd., Abingdon, United Kingdom (Cortisol, LTb4, Thromboxane, Endothelin-1, Substance P, c-AMP, and c-GMP); Ray Biotech Inc., Norcross, GA, USA (Leptin, EGF, Lipocalin, adiponectin, TNF alpha receptor 2 and HVEM); Sanbio B, Hycult biotech, Uden, The Netherlands (Calprotectin); Northwest Life Science Specialties, LLC, Vancouver, WA, USA (Isoprostane-2); Immundiagnostik GmbH, Bensheim, Germany (Zonulin); Cellmid Limited, Perth, Australia (Midkine); Diasource, Leuven, Belgium (Pregnenolone and vitamin D); Peninsula Laboratories, LLC, San Carlos, CA, USA (NPY); Promega Benelux BV, Leiden, The Netherlands (BDNF). LDN, Germany (Aldosterone); Hycult Biotech, USA (Nitrotyrosin).

All procedures were performed according to the manufacturer's instructions making use of an ELISA plate washer PW40 (Sanofi Pasteur). Read-outs of the Microtiter plate were digitally saved. Data were analyzed by making use of standard curves of OD values obtained by the Microtiter plate reader (Multiscan EF type 35, ThermoScientific) against (log transformed) concentrations as provided by the individual manufacturers of the kits. Individually measured patient sample values were obtained by linear interpolation of the sample OD value and the OD values of the standards. From each serum and urine sample creatinine levels were assessed and urine biomarker levels were corrected for the creatinine content. Patients and controls were only included with serum creatinine concentration within the normal range (excluding renal dysfunction).

2.3. Design of the study

A primary selection of biomarkers to be tested in serum and in first

morning urine was based on a thorough literature search in combination with a pilot study in 24 participants (12 MDD patients and their sex, age and ethnic matched healthy controls). The biomarkers included in this pilot cohort and their selection for the follow-up cohort is provided in supplementary Table S2. The selected biomarkers were subsequently tested in a cohort of 40 MDD patients and 47 healthy controls, fully independent of the pilot study. The results of this cohort were subsequently used for the design of the algorithm leading to the Bio-Depression Score (BDS, see below) and statistical validation by permutation analysis. Elimination of non-contributing biomarkers lead to a set of MDD associated biomarkers for which the association was confirmed by 5-fold cross validation.

2.4. Descriptive analysis

Descriptive statistics were calculated for the demographic parameters to describe the population. Numerical variables were summarized with means and standard deviations, while categorical variables were summarized with counts and percentages. Differences between the MDD and HC group were determined with Mann-Whitney *U* test for numerical data and Pearson's Chi-square test for categorical data. To determine median and variance differences in each biomarker for the MDD and HC group, the Mann-Whitney *U* test and Levene's test on heterogeneity were applied, respectively. These analyses were carried out using the Real Statistics add-in for Microsoft Excel.

2.5. Design of the algorithm and its validation to confirm association with MDD

The algorithm and its validation consist of 3 steps.

1. The algorithm (QBP) in which differences in the 'tails' of the distribution of multiple biomarkers between MDD and control subjects are combined into one single diagnostic score, the Bio-Depression Score (BDS):
 - a. For each biomarker separately the left and right tail dominance for the MDD and for the HC group is determined. *Left tail*: MDD dominates if the 10th percentile (P10) of the MDD group lies left of the P10 of the HC group, in which case the P1, P5 and P10 of the HC-distribution are used to define segments (see below). HC dominates if the order is opposite, in which case the P1, P5 and P10 of the MDD group are used to define the segments. *Right tail*: MDD dominates if the P90 of the MDD group lies to the right of the P90 of the HC group, in which case the P90, P95 and P99 of the HC-distribution are used to define the segments. HC dominates when this order is opposite, in which case the P90, P95 and P99 of the MDD group are used to define the segments.
 - b. For each biomarker the value-range is divided into segments according to the rules set out in 1.a.: values \leq P1; P1 < values \leq P5; P5 < values \leq P10; P10 < values < P90; P90 \leq values < P95; P95 \leq values < P99; values \geq P99.
 - c. Each biomarker value is normalized relative to these segments. In an MDD dominant tail the normalized values are assigned positive (i.e. a positive contribution to disease prediction), with value +1 for segments P5-P10 and P90-P95, value +2 for segments the P1-P5 and P95-P99, and value +3 for the segments < P1 and > P99. Similarly, in a HC dominant tail the normalized values are assigned negative (-1, -2, -3), respectively. For segment P10 - P90 the normalized value is assigned 0 (zero, i.e. no contribution to disease prediction).
 - d. A criterion for inclusion: in the left tail the presence of a minimum number of participants (percentage) in the dominant group at P10 and P5 and for the right tail at P90 and P95. A biomarker tail failing to reach the criterion is excluded: the normalized segment values in a failing tail are then set to value zero. Consequently, if both tails fail, the biomarker does not

contribute to disease prediction at all.

By application of the QBP algorithm each participant obtains per biomarker a normalized value ranging from -3 to +3. The BDS for a participant is then calculated as the sum of the normalized values of all contributing biomarkers. This generates per participant one single figure containing the cumulative information of all biomarkers. The higher the BDS the more likely the participant is associated with MDD. The lower the BDS the more likely the participant is not associated with MDD.

2. The second step describes the analysis of disease discrimination of the BDS followed by a permutation analysis to quantify the significance (*p*-value) of the discriminative power of the BDS using serum biomarkers only, urine biomarkers only, and a combination of serum and urine biomarkers. The generated BDS of each participant and its disease classification allows ROC analysis and calculation of the Area-Under-the-Curve (AUC_{Real}). MedCalc Statistical Software version 16.8 is used for comparison differences in AUC_{Real} between the groups 'serum only', 'urine only' and 'serum plus urine'. The higher AUC_{Real} , the better BDS discriminates between healthy and disease. The AUC_{Real} is used to determine the tail inclusion criteria mentioned in Step 1.d. above. After fixing all criteria in the algorithm, the discriminative power of the BDS is investigated with a permutation analysis. Then, the case/control indicator is randomly redistributed over the original biomarker data. Applying the algorithm to this random disease classification results in a BDS_{Random} and then applying ROC analysis generates an AUC_{Random} . Repeating this process at least 10,000 times an AUC_{Random} frequency distribution is generated. The proportion of random AUC_{Random} beyond the observed AUC_{Real} provides the *p*-value for the null hypothesis that the biomarkers are not associated with MDD.
3. The cross validation to support the association of the BDS with MDD for new samples. Discrimination does not necessarily provide good prediction of new samples. To determine the association between BDS and disease (MDD) a 5-fold cross-validation was performed. The validation was done on the biomarkers that were identified by the algorithm. The 87 participants were randomly divided into five parts (part 1: 18, part 2: 18, part 3: 17, part 4: 17 and part 5: 17 participants). No attempt was made to equally distribute MDDs and HCs, nor to match age and sex. Five separate sets were constructed: each contains 4 of the 5 subsets for training and the remaining subset for validation. For each set separately, the participants in the 'training subset' were used to determine the cut-offs for the percentiles P1, P5, P10, P90, P95 and P99 as described in Step 1. Next, based on these cut-offs, a BDS calculated for participants in the validation subset generates a ROC-curve for classification of the disease. This was executed for three situations: 'serum biomarkers only', 'urine biomarkers only', and 'for 'combined serum and urine biomarkers'. The results confirm the relevance for the association of the included biomarkers with MDD.

3. Results

3.1. Demographic characteristics

Table 1 shows the demographic characteristics of the subjects that were included in the 87-participant cohort. All subjects were of Caucasian origin. MDD subjects were matched for sex and age.

Control subjects had a median HAM-D17 score of 2.6 (range 0–8), while MDD subjects had a score of 19.0 (range 11–43). Due to insufficient amount of serum or identification errors, certain ELISAs were excluded in a minority of participants: 19 biomarkers were tested in all 87 participants, 16 biomarkers in all 86 participants, 2 biomarkers in 85 participants and 2 biomarkers in 83 participants. The results in serum are expressed as a concentration of the biomarker. The results in urine

Table 1
Demographic characteristics.

		All participants	Healthy Controls	MDD	Statistics
Sex	Male	35	19	16	Chi-square (2 × 2): P = .97
	Female	52	28	24	
Age (Yr)	Median	47.8	47.2	48.5	Mann-Whitney-U: P = .93
	SD	11.4	11.3	11.7	
Ethnicity	Caucasian	87	47	40	N.A.
HAM-D17	median	N.A.	2.6	19.0	Mann-Whitney-U: p < .001
	SD	N.A.	1.2	7.8	

Table 2
Mann-Whitney U test and Levene's test results.

	P-value			
	Serum		Urine	
	Mann-Whitney U test	Levene's test	Mann-Whitney U test	Levene's test
Adiponectin	–	–	0.35	0.67
Aldosterone	0.32	0.58	0.03 ^a	0.61
BDNF	0.91	0.10	–	–
Calprotectin	0.12	0.23	0.47	0.42
cAMP	0.55	0.53	0.38	0.56
cGMP	–	–	0.93	0.56
Cortisol	0.07	0.15	0.54	0.46
EGF	0.06	0.59	0.12	0.57
Endothelin	0.62	0.21	0.33	0.26
HVEM	0.41	0.72	0.69	0.62
Isoprostane	0.49	0.15	0.06	0.41
Leptin	0.90	0.53	0.66	0.96
Lipocalin	0.36	0.57	0.21	0.69
LTB4	0.08	0.06	0.91	0.05
Midkine	0.17	0.33	0.06	0.68
Nitrotyrosin	0.90	0.42	–	–
NPY	0.77	0.78	0.96	0.07
Pregnenolone	–	–	0.69	0.78
Substance P	0.08	0.08	0.13	0.36
Telomerase	0.30	0.52	–	–
Thromboxane B2	0.02 ^a	0.63	0.32	0.03 ^a
TNF R2	0.22	0.004 ^a	0.97	0.58
Vitamin D	0.90	0.18	–	–
Zonulin	0.37	0.002 ^a	–	–

^a Biomarkers with a p-value below p = .05.

are expressed as the ratio of biomarker to creatinine by dividing the biomarker concentration by that of creatinine. As a control for normal renal function, creatinine concentration was measured in serum as well and checked to remain within normal value ranges. Only those within the normal serum creatinine range were included.

Table 2 shows that differences at significant level in medians (Mann-Whitney U) were found in urine for Aldosterone and in serum for Thromboxane and in variances (Levene's) for Thromboxane in urine and for TNF and Zonulin in serum. Thus, these traditional statistical tests for location identified only a small number of relevant biomarkers.

3.2. BDS and association of biomarkers with MDD

The BDS was calculated for each of the 3 groups. Group 1 uses only the information of the 21 biomarker levels in serum. Group 2 uses only the information of the 19 biomarkers in urine. Group 3 uses all 40 biomarkers. The BDS was calculated according to the algorithm described in Step 1 of the Methods section.

The optimal limits for exclusion of non-performing tails are investigated (see Methods, Design of BDS, step 1(4)) by varying the limits for P10/P5 and P90/P95 from 0% to 40% and checking the effect on the AUC using all biomarkers (in Serum, in Urine and in serum plus urine). Without an exclusion limit, i.e. using the dominance classification only,

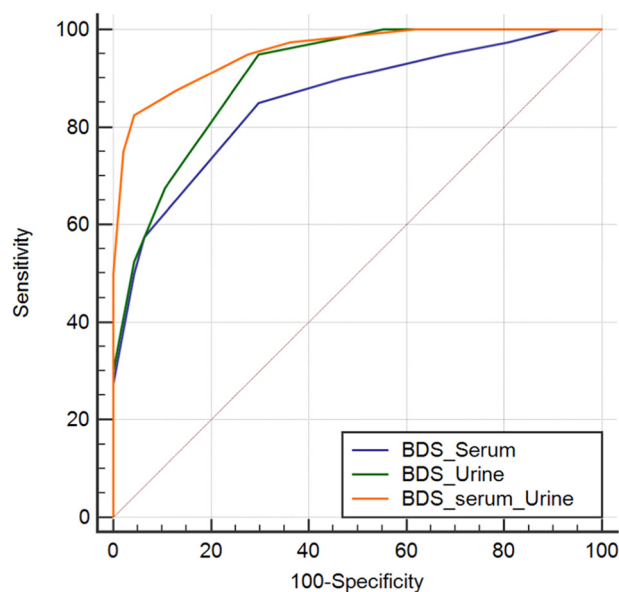


Fig. 1. ROC curves of the Bio Depression Score obtained for serum biomarkers, urine biomarkers, and serum plus urine biomarkers combined in the 87 participants. AUC = 0.853 for the 9 serum biomarkers only, AUC = 0.907 for 8 urine biomarkers only and 0.955 for the 17 combined serum + urine biomarkers. Comparison of ROC curves: 'Serum' versus 'Urine': P = .29; 'Serum' versus 'Serum plus Urine': P = .001; 'Urine' versus 'Serum plus Urine': p = .06.

the AUC becomes 0.876. At limit 40% for the P10 and P90, the AUC becomes 0.500 indicating that this limit excludes all biomarkers. The maximum AUC is obtained at 20% for P10 and P90 and at 17.5% for the P5 and P95 (for detailed information see Supplementary S3). These conditions exclude 23 biomarkers, leaving 9 biomarkers in Serum and 8 biomarkers in Urine with relevance for MDD. At higher exclusion percentages the AUC drops sharply. Thus, exclusion P10/90 at 20% and P5/95 at 17.5% is the condition generating the optimal result, which is chosen for further analysis. From high level of exclusion to lower levels, the biomarkers appeared in following order of decreasing relevance for serum: Thromboxane, TNF-R2, Cortisol, Substance P, BDNF, Calprotectin, cAMP, Zonulin and Leptin. In urine these were Thromboxane, Aldosterone, Isoprostane, Substance P, cGMP, HVEM, Midkine and Lipocalin. The ROC curves showing the results of the included biomarkers only are visualized in Fig. 1.

The permutation analyses using all 40 biomarkers and applying the 20%/17.5% inclusion criteria show significant BDS discrimination for biomarkers in urine (P < .001), in serum (P = .02) and combined in serum and urine (P < .001). The frequency distributions of the AUC_{Random} are visualized for serum, urine, and urine and serum respectively in Fig. 2. AUC_{Real} is positioned in the histogram as a vertical line.

3.3. Five-fold cross validation

By applying the 20%/17.5% exclusion limits, the contributing tails and dominances of the remaining 9 serum and 8 urine biomarkers are fixed. For each training-validation set, the percentile cut-off values are determined in the training subset and applied to the validation subset, leading to predicted AUC-values in the Validation subset (Table 3). The mean AUC of the ROC curves is lowest for the serum biomarkers, followed by the urine biomarkers and highest in the combined serum and urine biomarkers. Note that the SD of a validation subset is higher than in its training set in all cases. The validation sets show AUCs at the same level as the training set for serum and urine. For the combined serum plus urine the mean AUC is higher for both conditions. Thus, these

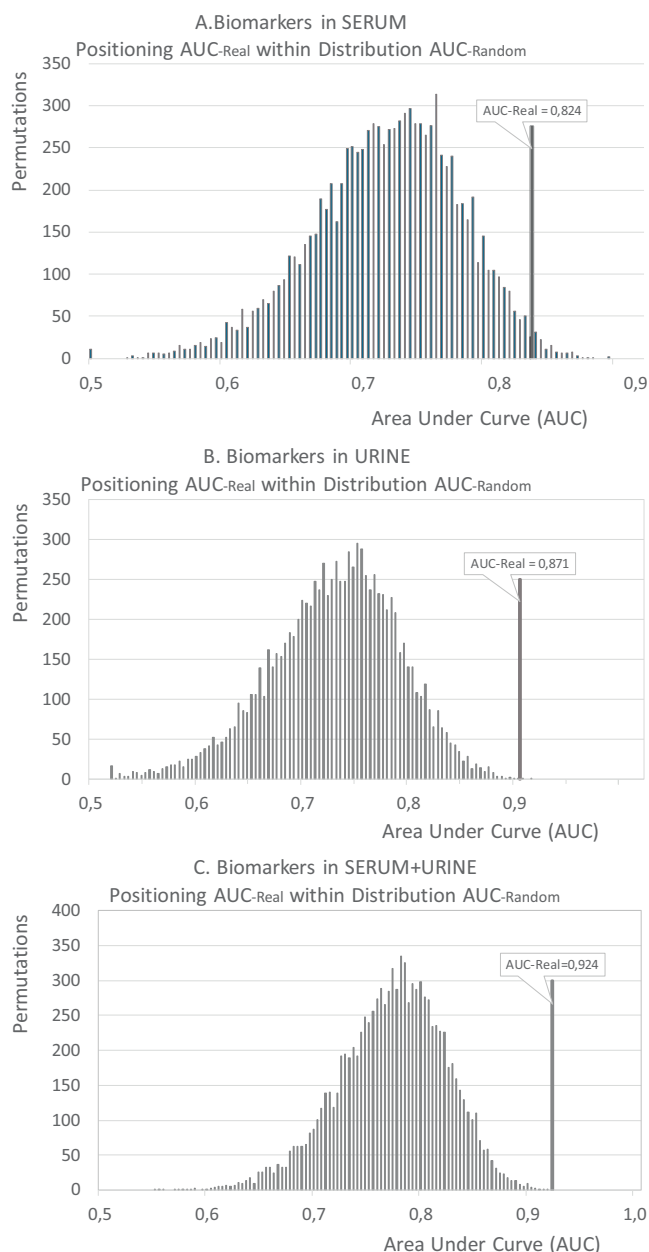


Fig. 2. Histogram of the case/control randomizations expressed as the number of randomization results at each AUC_{Random} found, relative to the position of AUC_{Real} . **Fig. 2A.** The histogram based on 10,001 permutations generated by the 21 serum biomarkers (mean $AUC_{Random} = 0.717 \pm 0.057$, fraction right of $AUC_{Real} = 1.74\%$, leading to a p-value of $p = .02$); **Fig. 2B.** The histogram of AUC_{Random} based on 10,004 permutations generated by the 19 Urine biomarkers (mean $AUC_{Random} = 0.709 \pm 0.057$, fraction right of $AUC_{Real} = 0.02\%$, leading to a significant p-value of $p < .001$); **Fig. 2C.** The histogram of AUC_{Random} based on 10,029 permutations generated by the combined 40 Serum plus Urine biomarkers (mean $AUC_{Random} = 0.796 \pm 0.046$, fraction right of $AUC_{Real} = 0.01\%$ leading to a significant p-value of $p < .001$). The difference in AUC_{Real} with the AUC presented in Fig. 1 reflect differences in tail optimization.

results fit well with the overall AUC results presented in Figs. 1 and 2, confirming the association between these biomarkers and MDD.

4. Discussion

The aim of this study was to identify biomarker sets that are measurably associated with MDD. Based on literature research we identified

six biochemical routes hypothetically associated with MDD. Within these hypotheses, 49 potential biomarkers were identified and tested in a pilot: 46 in serum and 44 in urine. Out of these we selected 40 biomarkers (21 in serum and 19 in urine) to be tested in the present case/control study. We have also included patients with a mild depression (HAMD-17 between 11 and 17): in the design of this study we anticipated that an association between biomarkers and MDD might become apparent also in the early or mild stages of depression. The background of the assumption is on the one hand that the disease phenotype would most probably not fully correlate with biochemical phenotype, and on the other hand that it may not be unlikely that early phases of disease also induce changes in biomarker homeostasis leading to adapted concentrations in either serum or urine. We found measurable associations with MDD using the combined effect of 9 biomarkers in serum, of 8 biomarkers in urine, and of 17 biomarkers in serum and urine combined. This result supports the assumptions made.

The AUCs found including only Caucasians in the analysis (this study) were higher than when 15 participants with other ethnicities were added to the analysis (not published). Some biomarkers were not identified in the combined ethnicities analysis whereas some did not appear in the analysis of the Caucasians only. This indicates that when investigating pathophysiology of biomarkers in psychiatric diseases ethnic background needs to be considered, irrespective what could have caused these differences.

Per participant only one sample has been taken without considering the variability that occurs within a participant. This association led to an Area Under the Curve of 0.955 in a Receiver Operating Characteristic analysis and was confirmed by 5-fold cross validation. The large value of the AUC may suggest that inter-individual variability dominates the intra-individual variability within biomarkers, but we recommend that future confirmation studies would apply repeated samples to be able to estimate the index of individuality and provide a view of the prediction within individuals. Biomarkers in serum appeared somewhat less efficient than those in urine. The combined result in serum plus urine was better than either serum or urine. For all hypotheses [3] for pathophysiology leading to MDD measurable biomarkers were found: Cortisol, Aldosterone and Substance P for abnormal HPA-axis functioning; Substance P for monoaminergic changes; TNF-R2, HVEM, Lipocalin and Substance P for altered immune system functioning and inflammation, BDNF, Midkine and TNF-R2 for changes in neurogenesis/neuroplasticity, Endothelin, Thromboxane-B2, Zonulin, cAMP, cGMP, Leptin and F2-isoprostane for oxidative stress/endothelial dysfunction and Thromboxane-B2, Aldosterone and Substance P for altered magnesium metabolism and mineral homeostasis.

We realize that our proposed method for biomarker selection and disease prediction (QBP approach, to be published) is different from commonly used statistical or machine learning approaches. Our focus has been on biomarker selection, which would rule out certain approaches that require all (linearly transformed) biomarkers in their prediction of the disease. The relatively low sample size also makes certain machine learning approaches less applicable (e.g. support vector machine). Furthermore, the QBP approach has been intensively studied under all kinds of settings using simulation studies where the approach is compared to more common approaches (e.g. discriminant analysis; logistic regression). When differences in the probability distributions between cases and controls are determined by variation in biomarkers, the QBP method is superior to the commonly known approaches, and it is equivalent to the commonly known approaches when differences are due to location. An advantage of the QBP is that it is conservative in selecting biomarkers, indicating that we would not eliminate important biomarkers.

A total of 17 biomarkers were identified, whereas the classical statistical differences analyzed by the non-parametric Levene's test and Mann-Whitney U test detected 5 biomarkers only. An alternative analytic approach to measure the combined effects of the biomarkers could have been machine learning. Our approach is of a basic simplicity and

Table 3
Results of five-fold cross validation expressed in the AUC's of the subsets.

Data set	5-fold cross validation (AUC)					
	Serum 9 biomarkers		Urine 8 biomarkers		Serum + Urine 17 biomarkers	
	Training	Validation	Training	Validation	Training	Validation
Sub set 1	0.827	0.733	0.835	0.933	0.895	0.847
Sub set 2	0.780	0.861	0.856	0.868	0.959	0.931
Sub set 3	0.809	0.764	0.867	0.840	0.922	0.813
Sub set 4	0.799	0.846	0.847	0.923	0.912	0.965
Sub set 5	0.794	0.799	0.884	0.778	0.921	0.850
Mean	0.802	0.801	0.858	0.869	0.922	0.881
± SD	0.016	0.048	0.017	0.057	0.021	0.057

The means and SDs of the of the combined subsets are given in bold to differentiate them from the individual results of each subset.

generally applicable: it cumulates the different biomarker outcomes per individual into a one-figure outcome. It can be used to associate biomarker outcomes with clinical disease and relate to pathophysiology.

So far, only a handful of studies have combined results of multiple biomarkers into one single diagnostic test [10–15]. A main advantage of the approach used in this study as compared to other depression-related biomarker studies is that we are the first to use a method that includes differences in variation and distribution between MDD and control groups, thus optimizing the use of available data.

Furthermore, with the exception of Zheng et al. [15], most studies that have investigated biomarkers for depression focus on serum biomarkers. We have chosen to investigate biomarker levels in both serum and urine in this study, and we have shown that urine biomarkers may be of added value for the identification of biomarkers: in urine we found unique biomarkers such as HVEM, Midkine and cGMP. One of the reasons for this might be that serum levels may show a large variation over the day, limiting their potential use. Biomarker levels in urine are a collection over a time period and may in such cases give more consistent results. The use of urine biomarkers has several advantages over the use of serum biomarkers. It is easy to collect and does not require the use of invasive techniques. In addition, it is relatively easy to test for proteins in urine, as urine is relatively protein-poor and thus the chance of cross-reactivity is smaller than in serum.

It would be of interest to investigate whether specific biomarkers are related to any specific phenotype of MDD, such as for example treatment-resistant depression, the euthymic state (trait), severity of depression (state), melancholia, atypical depression, or depression with anxiety features. The present study does not allow to draw conclusions at this level, as it was designed to show a primary relation of biomarkers and depression and further research is needed to show relation with phenotype subtyping.

This work is intended to investigate measurable biomarkers levels associated with MDD. Our working hypothesis assumes that apparent clinical MDD causes measurably differences in biomarker levels from controls, irrespective status of medication. Nevertheless, medication and the duration thereof may induce measurable effects on biomarker levels [16–19]. Several studies have demonstrated that levels of biomarkers may be influenced also in the absence of treatment response. Turck et al. [17] for example have analyzed plasma proteins at baseline and after 6 weeks of treatment and found 46 proteins to be altered at 6 weeks in non-responders against 43 proteins that were altered in responders. Another study found a decrease of several plasma cytokines (IL-2, IL-4, IL-6, IL-10 and IFN- γ) in both responders and non-responders to either Sertraline treatment, a course of transcranial Direct Current Stimulation (tDCS) or Sertraline and tDCS combined [20]. The effect of treatment on levels of biomarkers in responders/remitters vs. non-responders/non-remitters may be dependent upon the chosen treatment strategy. For example, Chen et al. [19] demonstrated an increase in levels of cytokines IL-6, IFN- γ and TNF- α in non-remitters and a decrease in levels of cytokines IL-2, IL-4 and IL-5 in remitters after

8 weeks treatment with paroxetine. Venlafaxine treatment however resulted in a decrease in levels of cytokines IL-2, TNF- α , IL-4, IL-5, IL-10 and IL-1 β both in non-remitters and remitters and a decrease in IFN- γ only in non-remitters. Another study demonstrated opposite effects of Amitriptyline and Paroxetine on BDNF levels [21]. Furthermore, a change in the levels of biomarkers may in some cases precede clinical improvement [22]. Thus, despite the depressed state of all patients included in this study, the medication status of part of the patients may have influenced the levels of some of the biomarkers. Because of the limited group size and the variation in types and duration of medication we have not attempted to differentiate between these. Differentiating status of medication is considered an essential step for future investigation requiring larger dedicated patient cohorts.

Of the previously performed studies investigating combinations of biomarkers for depression, the most consistent results have been obtained by Papakostas et al. [13] and Bilello et al. [14]. Papakostas and colleagues show that a previously identified panel of 9 biomarkers, the results of which were combined into an “MDD score”, can diagnose depression with a sensitivity of > 90% and a specificity of > 80% [13]. These results have been validated in a separate cohort, with the addition of two additional factors to the algorithm calculating the MDD score: gender and body mass index (BMI), and correction for circadian serum cortisol levels [14]. We checked the effect of correction of circadian cortisol for the cortisol samples from which a sampling time-point was available and found that this correction induced a favorable effect of the AUC from 0.955 to 0.970 (data not shown), thus further confirming both serum and urine as body fluids providing independently biomarker results. Our overall result for biomarkers in serum and urine combined is similar to that of Papakostas et al. Future research would need to determine if additional factors such as BMI and gender may further improve this result.

Together with these studies, our study demonstrates that the use of biomarker panels is a promising method to investigate MDD pathophysiology at biomarker level. The study of combinations of biomarkers originating from different “hypotheses” may better elucidate the heterogeneous pathophysiology of depression, and may therefore increase the knowledge about the relation between biomarkers and disease state [3]. In addition, preliminary evidence indicates that the combination of biological markers, clinical variables and self-report might enable physicians to predict the development of a subsequent episode of MDD in patients suffering from other psychiatric disorders, which may facilitate early intervention by timely starting antidepressant treatment [23].

With respect to this case/control study, several limitations should be kept in mind. We used a relatively low depression severity cut-off (HAM-D \geq 11) and we included a wide range of HAM-D scores (ranging from HAM-D = 11 to HAM-D = 43, average HAM-D = 19). The inclusion of patients with mild-to-moderate depression might have limited the level of association found between the MDD and control groups. Nevertheless, we still found a significant discriminating effect, leading

to quite high AUC levels, indicating that the biomarkers might also be measurable in subjects with mild-to-moderate depression. Second, although the phenotype permutation analysis has shown that the resulting AUC cannot be explained by “overfitting” and the 5-fold cross validation has shown, that the results have predictive potency, these results should still be confirmed in a separate and independent cohort of subjects. Third, although this study indicates that this panel of biomarkers can successfully discriminate between MDD subjects and healthy controls, we have used quite stringent exclusion criteria for comorbidity in order to increase homogeneity between the MDD and control groups on areas other than depression score. Thus, the clinically relevant question whether this biomarker panel can also discriminate between MDD and other psychiatric or somatic disorders needs to be investigated in follow-up studies (MDD specificity). Fourth, the effects of confounders are to be addressed: gender, ethnicity, age, body mass index, smoking and other differences may exist. Fifth, 19 of the 40 MDD patients in this study were on antidepressant medication, which might have influenced levels of certain biomarkers. In this study, the clinical presentation as judged by the Hamilton score was taken as reference: patients presenting with MDD symptoms, without reconsidering antidepressant use, leading to the results obtained. Follow-up studies are needed to elucidate any differences related to antidepressant use. Despite these limitations, this study provides evidence towards the elucidation of heterogenic pathophysiology of MDD at the level of biomarkers.

In addition, with the elucidation of the pathophysiology, biomarkers may become available for efficient diagnosis, they could potentially be used for the identification of those at risk for developing MDD and the prediction of treatment response and the identification of MDD subtypes. The value of these biomarkers in combination with the BDS algorithm with regard to such application needs to be determined in future studies. As this panel contains biomarkers related to important pathophysiological mechanisms related to certain MDD subtypes [24–28] it is a promising perspective that MDD subtypes could be identified based on the patient's specific biomarker scores.

5. Conclusion

We have demonstrated that a panel of biomarkers related to different aspects of MDD pathophysiology can measurably be associated to MDD subjects using both serum and urine as body fluid matrix. Although future research is needed to confirm these results in a separate and larger group of subjects, and optimize the biomarkers association by including confounders, medication, sex, age BMI, ethnicity, et cetera, this study may be a first step towards the development of a biomarker-based test for depression that can be used in a clinical setting. Particularly non-psychiatric physicians, such as general practitioners, could benefit.

Declaration of Competing Interest

Meddens and Arnoldussen hold stocks in Brainlabs. The literature search by Bosker and Gladkevich in the early phase of the PIDON project was partially financed by Brainlabs. Bosker was at the time of this study scientific director of Brainlabs but had no financial relation with the company. Brainlabs has filed separate patents covering the diagnostic use of the biomarkers as described in this manuscript and the statistical methodology applied. Ownership of these patents has been transferred to Brainscan B.V. Meddens and Arnoldussen hold stocks in Brainscan B.V. and are currently employed at Brainscan B.V. Other authors declare no potential conflicts of interest.

Acknowledgements

This work is based on a grant called PIDON originating from Dutch governmental institutions (provinces Gelderland and Overijssel,

registered under grant number PID_09_02/120329), in which Brainlabs, University of Groningen and University Medical Center Groningen participated. The PIDON grant covered the activities of literature search, study design, patient inclusions, laboratory analysis and data reduction.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpsychores.2019.109796>.

References

- [1] E. Bromet, L.H. Andrade, I. Hwang, N.A. Sampson, J. Alonso, G. de Girolamo, et al., Cross-national epidemiology of DSM-IV major depressive episode, *BMC Med.* 9 (2011 Jul 26) 90.
- [2] P. Sobocki, B. Jönsson, J. Angst, C. Rehnberg, Cost of depression in Europe, *J. Ment. Health Policy Econ.* 9 (2) (2006 Jun) 87–98.
- [3] M.C. Jentsch, E.M. Van Buel, F.J. Bosker, A.V. Gladkevich, H.C. Klein, R.C. Oude Voshaar, et al., Biomarker approaches in major depressive disorder evaluated in the context of current hypotheses, *Biomark. Med.* 9 (3) (2015) 277–297.
- [4] M. Maes, P. Galecki, Y.S. Chang, M. Berk, A review on the oxidative and nitrosative stress (O&NS) pathways in major depression and their possible contribution to the (neuro)degenerative processes in that illness, *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* 35 (3) (2011 Apr 29) 676–692.
- [5] A. Halaris, Inflammation, heart disease, and depression, *Curr. Psychiatry Rep.* 15 (10) (2013 Oct) 400.
- [6] A. Serefko, A. Szopa, P. Wlaź, G. Nowak, M. Radziwoń-Zaleska, M. Skalski, et al., Magnesium in depression, *Pharmacol. Rep.* 65 (3) (2013) 547–554.
- [7] G.A. Eby, K.L. Eby, Magnesium for treatment-resistant depression: a review and hypothesis, *Med. Hypotheses* 74 (4) (2010 Apr) 649–660.
- [8] Biomarkers Definitions Working Group, Biomarkers and surrogate endpoints: preferred definitions and conceptual framework, *Clin. Pharmacol. Ther.* 69 (3) (2001 Mar) 89–95.
- [9] M.H. Paus, van den Heuvel, R. Edwin, M.J. Meddens, Binary disease prediction using tail quantiles of the distribution of continuous biomarkers, (2019) (Submitted).
- [10] E. Domenici, D.R. Willé, F. Tozzi, I. Prokopenko, S. Miller, A. McKeown, et al., Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections, *PLoS One* 5 (2) (2010 Feb 11) e9166.
- [11] S.E. Arnold, S.X. Xie, Y. Leung, L. Wang, M.A. Kling, X. Han, et al., Plasma biomarkers of depressive symptoms in older adults, *Transl. Psychiatry* 2 (2012 Jan 03) e65.
- [12] K. Alawam, E. Dudley, R. Donev, J. Thome, Protein and peptide profiling as a tool for biomarker discovery in depression, *Electrophoresis* 33 (24) (2012 Dec) 3830–3834.
- [13] G.I. Papakostas, R.C. Shelton, G. Kinrys, M.E. Henry, B.R. Bakow, S.H. Lipkin, et al., Assessment of a multi-assay, serum-based biological diagnostic test for major depressive disorder: a pilot and replication study, *Mol. Psychiatry* 18 (3) (2013 Mar) 332–339.
- [14] J.A. Bilello, L.M. Thurmond, K.M. Smith, B. Pi, R. Rubin, S.M. Wright, et al., MDDScore: confirmation of a blood test to aid in the diagnosis of major depressive disorder, *J. Clin. Psychiatry* 76 (2) (2015 Feb) 199.
- [15] P. Zheng, Y. Wang, L. Chen, D. Yang, H. Meng, D. Zhou, et al., Identification and validation of urinary metabolite biomarkers for major depressive disorder, *Mol. Cell. Proteomics* 12 (1) (2013 Jan) 207–214.
- [16] G.I. Papakostas, Surrogate markers of treatment outcome in major depressive disorder, *Int. J. Neuropsychopharmacol.* 15 (6) (2012 Jul) 841–854.
- [17] C.W. Turck, P.C. Guest, G. Maccarrone, M. Ising, S. Kloiber, S. Lucae, et al., Proteomic differences in blood plasma associated with antidepressant treatment response, *Front. Mol. Neurosci.* 10 (2017 Aug 31) 272.
- [18] C.A. Kohler, T.H. Freitas, B. Stubbs, M. Maes, M. Solmi, N. Veronese, et al., Peripheral alterations in cytokine and chemokine levels after antidepressant drug treatment for major depressive disorder: systematic review and meta-analysis, *Mol. Neurobiol.* 55 (5) (2018 May) 4195–4206.
- [19] C.Y. Chen, Y.W. Yeh, S.C. Kuo, C.S. Liang, P.S. Ho, C.C. Huang, et al., Differences in immunomodulatory properties between venlafaxine and paroxetine in patients with major depressive disorder, *Psychoneuroendocrinology* 87 (2018 Jan) 108–118.
- [20] A.R. Brunoni, R. Machado-Vieira, C.A. Zarate, L. Valiengo, E.L. Vieira, I.M. Bensenor, et al., Cytokines plasma levels during antidepressant treatment with sertraline and transcranial direct current stimulation (tDCS): results from a factorial, randomized, controlled trial, *Psychopharmacology* 231 (7) (2014 Apr) 1315–1323.
- [21] R. Hellweg, A. Ziegenhorn, I. Heuser, M. Deuschle, Serum concentrations of nerve growth factor and brain-derived neurotrophic factor in depressed patients before and after antidepressant treatment, *Pharmacopsychiatry* 41 (2) (2008 Mar) 66–71.
- [22] H. Murck, M.C. Braunsch, C. Konrad, D. Jezova, T. Kircher, Markers of mineralocorticoid receptor function: changes over time and relationship to response in patients with major depression, *Int. Clin. Psychopharmacol.* 34 (1) (2019 Jan) 18–26.
- [23] M.G. Gottschalk, J.D. Cooper, M.K. Chan, M. Bot, B.W.J.H. Penninx, S. Bahn, Discovery of serum biomarkers predicting development of a subsequent depressive

- episode in social anxiety disorder, *Brain Behav. Immun.* 48 (2015 Aug) 123–131.
- [24] F. Lamers, N. Vogelzangs, K.R. Merikangas, P. de Jonge, A.T.F. Beekman, B.W.J.H. Penninx, Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression, *Mol. Psychiatry* 18 (6) (2013 Jun) 692–699.
- [25] H.E. Duijvis, N. Vogelzangs, N. Kupper, P. de Jonge, Brenda W.J.H. Penninx, Differential association of somatic and cognitive symptoms of depression and anxiety with inflammation: findings from the Netherlands study of depression and anxiety (NESDA), *Psychoneuroendocrinology* 38 (9) (2013 Sep) 1573–1585.
- [26] K. Becking, L. Boschloo, N. Vogelzangs, B.C.M. Haarman, R. Riemersma-van der Lek, B.W.J.H. Penninx, et al., The association between immune activation and manic symptoms in patients with a depressive disorder, *Transl. Psychiatry* 3 (2013 Oct 22) e314.
- [27] C. Stetler, G.E. Miller, Depression and hypothalamic-pituitary-adrenal activation: a quantitative summary of four decades of research, *Psychosom. Med.* 73 (2) (2011 Feb-Mar) 114–126.
- [28] N. Vogelzangs, H.E. Duijvis, A.T.F. Beekman, C. Kluit, J. Neuteboom, W. Hoogendijk, et al., Association of depressive disorders, depression characteristics and antidepressant medication with inflammation, *Transl. Psychiatry* 2 (2012 Feb 21) e79.