Towards a neurobiological view of depression

van Buel, Erin

Major Depressive Disorder is associated with changes in a cluster of serum and urine biomarkers; implications for the development of a biomarker-based diagnostic test

3

E.M. van Buel[1,3*], M.J.M. Meddens[2*], A.E. Arnoldussen[2], F.J. Bosker[3*], C. le Conge[2], E.R. van den Heuvel[4], W. Bohlmeijer[5], J.A. de Boer[3], I. Kema[3], A. Muller Kobold[3], L.J.M. Boonman-de Winter[6], A. van Erven[7], B. Timmers[8], P. Maas[9], R. Korpershoek[10], T. Veerman[11], J.S. Kamphuis[12], A.V. Gladkevich[3], R.A. Schoevers[3], P.G.M. Luiten[1], U.L.M. Eisel[1], H.C. Klein[3]

*These authors contributed equally to this manuscript.

Manuscript in preparation.

Abstract

Major Depressive Disorder (MDD) is a heterogenous disorder with a large symptomatic overlap with other psychiatric and somatic disorders. As a result, diagnosis may be complicated, particularly for the non-psychiatrist physician. This study aims at providing a practical clinical test to assist in the diagnosis of MDD by using a small set of serum and/or urine biomarkers. To this end, we analyzed urine and serum samples of 51 MDD patients and 51 age-, sex-, and ethnicity-matched controls for levels of 40 potential MDD biomarkers (21 serum biomarkers and 19 urine biomarkers). We developed an algorithm to select biomarkers on the basis of differences in variation and distribution between groups, combine the outcome of the selected biomarkers, and calculate a depression probability score for clinical discrimination of depressed from euthymic subjects (the "bio depression score"). Based on this algorithm, 11 urine biomarkers and 6 serum biomarkers were selected that together reached an area under the curve (AUC) of 0.889 in ROC analysis. Inclusion of only urine biomarkers or only serum biomarkers resulted in an AUC of 0.859 and 0.755 respectively. A phenotype permutation analysis showed a significant discrimination between MDD and control subjects for biomarkers in urine ($P<0.001$) and for the biomarkers in serum and urine combined ($P<0.001$), but not for serum biomarkers only ($P=0.20$). An internal cross-validation confirmed the predictive value of this set of biomarkers. Although awaiting confirmation in a separate cohort, these results are an important step towards the development of a biomarker-based depression test.

Introduction

Major depressive disorder (MDD), with a lifetime prevalence of around 15%, is a major cause of disability in the western world [1,2]. Diagnosis is currently based on symptomatic criteria. However, due to the heterogeneous nature of MDD and symptomatic overlap with other psychiatric and somatic disorders, diagnosis may be complicated [3]. Especially non-psychiatrist physicians may have difficulties recognizing MDD [4,5] and could benefit from diagnostic methods that can be used in addition to existing methods. This study describes the development of a biomarker-based test for depression that can be used by non-psychiatrist physicians as an easy and quick method to determine the likelihood and possibly also the diagnosis of MDD.

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" [6]. 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. TNFα, IL-1β), neurotrophic factors (e.g. BDNF, VEGF), and hormones (e.g. cortisol). However, none of these biomarkers reaches sufficient sensitivity and specificity on its own [3]. This may be in part due to the complicated pathophysiology underlying 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, and 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. Possibly, a combination of biomarkers reflecting divergent dysfunctions in MDD might prove more successful [3].

In a novel method to discriminate between patients and controls, we eventually selected a subset of biomarkers, covering all important hypotheses of MDD pathophysiology: 1) the HPA-axis hypothesis, 2) the monoamine hypothesis, 3) the immune-inflammation hypothesis, 4) the neurogenesis and neuroplasticity hypothesis, 5) the oxidative stress and endothelial dysfunction hypothesis and 6) the magnesium deficiency/ion binding hypothesis. For a description of the first four hypotheses and their relation to potential biomarkers for MDD, we refer to Jentsch et al. [3]. The role of oxidative stress and endothelial dysfunction in MDD is extensively reviewed by Maes and colleagues [7] and Halaris [8]. Serefko *et al*. [9] and Eby & Eby [10] have reviewed the potential association between MDD and magnesium deficiency. A summary of the scientific support for these hypotheses and a list of biomarkers included in this study that are related to these hypotheses are provided in Table 1.

Table 1 *(continues on next page). Hypotheses for the pathophysiology of Major Depressive Disorder.*

Hypothesis	**Description**	**Typical biomarkers**
HPA-axis / stress-related	There is a large body of evidence supporting an association between MDD and functional disturbances of the hypothalamus-pituitary-adrenal (HPA) axis, the main stress system of the body [11-13].	Cortisol, aldosterone, NPY, pregnenolone, substance P
Monoamine	The monoamine hypothesis suggests that abnormalities of monoaminergic systems (serotonin, norepinephrine and dopamine) may be associated with MDD. This theory is supported by findings of impaired monoaminergic neurotransmission and decreased levels of monoamine neurotransmitters and their metabolites in brain tissue, cerebrospinal fluid, and blood of MDD patients [14-21].	NPY, substance P, BH4
Immune-inflammation	The immune-inflammation hypothesis of depression is supported by evidence that MDD is associated with pro-inflammatory changes, including elevated levels of pro-inflammatory cytokines, decreased levels of anti-inflammatory cytokines and activation of microglia as well as peripheral immune cells. In addition, pro-inflammatory cytokines have been shown to induce depression-like behavior in animals [22-27].	TNF-R2, HVEM, lipocalin, calprotectin, substance P, vitamin D

Table 1 (continuation).

Hypothesis	Description	Typical biomarkers
Neurogenesis / neuroplasticity	MRI studies have demonstrated decreased hippocampal volumes in MDD patients, suggesting decreased neurogenesis. In addition, expression of growth factors may be disturbed in depressed patients. Animal studies further support an association between decreased neurogenesis /neuroplasticity and depression-like behavior [28-30].	BDNF, VEGF, EGF, PEDF, Midkine, TNF-R2
Oxidative stress /endothelial dysfunction	Endothelial dysfunction results in an imbalance in the production of vasodilating and vasoconstricting substances and may be triggered by pro-inflammatory cytokines. Endothelial dysfunction is believed to play an important role in the association between MDD and myocardial. Endothelial dysfunction may lead to oxidative stress, resulting in increased production of reactive oxygen species (ROS) and induction of inflammatory processes [7,8,31,32].	Endothelin-1, adiponectin, thromboxane-B2, Zonulin, nitrotyrosine, cAMP, cGMP, LTB4, Leptin, F2-isoprostane
Magnesium metabolism / mineral homeostasis	An increasing body of evidence implicates magnesium (Mg^{2+}) deficiency in the pathophysiology of MDD (for review, see. Animal studies have demonstrated that magnesium malnutrition induces depressive-like behavior and anxiety. In addition, there are indications that magnesium supplementation can reduce symptoms of depression in rodents and humans [10,33-38].	Thromboxane-B2, aldosterone, substance P, thromboxane-B2

A major problem in biomarker-based diagnostics 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 taking into account that each biomarker obeys different variance rules between cases and controls.

In this manuscript we perform an analysis that prevents these pitfalls. The tool is used to select those biomarkers that "behave" different between cases and controls without 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 validity of this subset was tested. We describe an algorithm to select biomarkers, combine the outcome of the selected biomarkers, and calculate a depression probability score for clinical discrimination of depressed from euthymic subjects. The tool provides for each subject a biometric depression score, tentatively named BDS (Bio-Depression-Score). Aim of the study is to provide a practical clinical test to assist in the diagnosis of major depression by using a small set of serum and/or urine biomarkers.

Methods

Patients

MDD patients were recruited in cooperation with general practitioners, psychiatric clinics and through advertisements in local and national newspapers. Inclusion criteria included: age 18-65, fulfilled DSM-IV criteria for unipolar MDD, a HAM-D score higher than 10 and informed consent. Exclusion criteria included presence of another primary psychiatric disorder, alcohol or substance use disorders, inflammatory or systemic diseases, metabolic disorders or other disorders that might affect mood, and pregnancy. Patients with (n=30) and without (n=21) anti-depressive medication were included. Healthy controls (HC) were recruited via general practitioners and advertisements in local and national newspapers. Healthy controls had to be free of any major axis I diagnosis and were matched for gender, age and ethnicity.

The 17-item Hamilton Depression Rating Scale (HAM-D) was used to assess symptoms of depression. In addition, a Mini International Neuropsychiatric Interview (MINI) was conducted. A researcher who was trained in the use of these questionnaires executed all questionnaires, and the final MDD diagnosis was made by a psychiatrist.

Participants were asked to deliver 50 ml of blood through venipuncture as well as 50 ml of first morning urine. Blood was collected in serum separation tubes, allowed to clot and centrifuged at 3000 x g for 10 minutes. Serum supernatant was divided into aliquots and stored at -80°C. Urine samples were centrifuged for 10 minutes at 1000 x g to precipitate any particles and cells; the supernatant was collected, divided into aliquots and stored at -80°C.

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, TNFalpha 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 Microtiterplate 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 standard. From each serum and urine sample creatinin was tested and urine biomarker results were expressed relative to the creatinine content. Patients and controls were only included with serum creatinine concentration within the normal range (excluding renal dysfunction).

Table 2 (continues on next page). Biomarker selection in the pilot and case/control study. X: selected biomarkers.

Biomarkers	Biomarkers tested in			
	2 x12 Pilot		2 x 51 Case/Control	
	Serum	Urine	Serum	Urine
Adiponectin	X	X	-	X
Aldosteron	X	X	X	X
Bcl-2	X	X	-	-
BDNF	X	-	X	-
Beclin-1	X	-	-	-
Bfgf	X	X	-	-
BH4	X	X	-	-
Biopterin	-	X	-	-
Calprotectin	X	X	X	X
Calreticulin	X	X	-	-
CAMK2B	X	X	-	-
Camp	X	X	X	X
CCK	X	X	-	-
Cgmp	X	X	-	X
Cortisol	-	X	X	X

Biomarkers	Biomarkers tested in			
	2 x12 Pilot		2 x 51 Case/Control	
	Serum	Urine	Serum	Urine
Digoxine	X	X	-	-
EGF	X	X	X	X
Endothelin	X	X	X	X
F2-Isoprostane	X	-	-	-
GABA	X	X	-	-
Galectin-8	X	X	-	-
HVEM	X	X	X	X
IGF1	X	X	-	-
IL-6	X	X	-	-
Isoprostane	-	X	X	X
Leptin	X	X	X	X
Lipocalin	X	X	X	X
LOX1	X	X	-	-
LTB4	X	X	X	X
MBP	X	X	-	-
Midkine	X	X	X	X
MMP-1	X	-	-	-
Myeloperoxidase	X	X	-	-
Neopterin	X	X	-	X
Neuropeptide Y	X	X	X	X
NGF	X	X	-	-
Nitrotyrosine	X	X	X	-
PEDF	X	X	-	-
PLA2G7-PAF	X	-	-	-
Pregnenolon	X	X	X	X
Prostaglandin E2	X	X	-	-
Substance P	X	X	X	X
Svegf	X	X	-	-
Thromboxane B2	X	X	X	X
TNF-R2	X	X	X	X
Vasopressin	X	X	-	-
VILIP	X	X	-	-
Vit-D	X	X	X	-
Zonulin	X	X	X	-

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 included biomarkers in this pilot cohort and the selection for the follow-up cohort is provided in Table 2. The selected biomarkers were subsequently tested in a cohort of 51 MDD patients and 51 matched healthy controls. The results of this cohort were subsequently used for the design of an algorithm leading to the diagnostic score (BDS, see below) and statistical validation by permutation analysis. After elimination of non-contributing biomarkers the predictive value of the diagnostics score was investigated by 5-fold cross validation.

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 all performed with SPSS statistical software, version 23.

Design of the algorithm to predict MDD

The first step describes the algorithm that combines the left and right tail measurements of multiple biomarkers into one single diagnostic score, the Bio Depression Score (BDS). In the second step, the performance measure for discrimination of the BDS is described together with permutation test to quantify the significance (p-value) of the discriminative power of the BDS in serum only, in urine only, and in serum and urine combined. In the third and final step, a cross validation is described to quantify the predictive value of the BDS for new samples.

Step 1: The Bio Depression Score

1. For the left and right tail of each biomarker separately, the dominance of the MDD and HC group is determined first:
   - **a.** Left tail. MDD dominates if the 10th percentile of the MDD group lies left of the 10th percentile of the HC group. HC dominates if the order of these 10th percentiles is opposite. If the MDD-group dominates, the 1st (P1), 5th (P5) and 10th (P10) percentiles of the HC-distribution are used as cut-offs to form scores in the BDS and vice versa.
   - **b.** Right tail. MDD dominates if the 90th percentile of the MDD group lies to the right of the 90th percentile of the HC group. HC dominates when this order is opposite. If the MDD group dominates, the 90th (P90), 95th (P95) and 99th (P99) percentiles of the HC-distribution are used as cut- offs to form scores in the BDS, and vice versa.
2. Based on the above cut-offs, the distribution of each biomarker is divided into 7 segments: values ≤P1; P1< values ≤P5; P5< values ≤P10; P10<values<P90; P90≤values<P95; P95≤values<P99; values ≥P99.

3. Each biomarker value is transformed into a score. Within the segment P10 to P90 the score will become zero (no contribution to disease prediction). In a MDD dominant tail the scores are assigned positive (+1, +2, +3) values, with higher numbers for segments further away in the tails. In a HC dominant tail the score is assigned negative (-1, -2, -3) scores, with lower values for segments further away in the tail.
4. Dominance alone is not enough. To include a tail for a biomarker, the left (or right) tail of the dominant group should be substantially further to the left (or right) than the non-dominant tail. For instance, criteria may define that the mass of the left (or right) tail of the dominant group should be at least 20% at the cut-off P10 (P90) of the non-dominant group. In addition, more than 15% of the dominant group should be left (right) of the cut-off P5 (P95). If one biomarker tail does not reach these two pre-set tail criteria, the BDS score for participants with biomarker values in these tails are set to zero. If both tails of a biomarker fail to satisfy these tail criteria, the biomarker does not contribute to disease prediction.
5. The above described algorithm is summarized in Table 3.
6. By application of the above algorithm each participant obtains per biomarker a score ranging from -3 to +3. The BDS for a participant is the sum of the scores for all biomarkers. Thus the BDS is the cumulative information from all incorporated biomarkers towards the presence or absence of MDD. A positive score indicates an increased likelihood of MDD, while a negative score indicates a decreased likelihood of MDD. The higher the score the more likely the disease is present and the lower the score the more likely the disease is not present.

***Table 3**. Summary of assigned normalizing values in all options*

Dominance in tail			Scores in the different segments						
	Left	Right	≤P1	>P1 - ≤P5	>P5 - ≤P10	>P10 - <P90	≥P90 - <P95	≥P95 - <P99	≥P99
Both Tails	HC	MDD	-3	-2	-1	0	1	2	3
	HC	HC	-3	-2	-1	0	-1	-2	-3
	MDD	HC	3	2	1	0	-1	-2	-3
	MDD	MDD	3	2	1	0	1	2	3
One Tail	HC	--	-3	-2	-1	0	0	0	0
	--	HC	0	0	0	0	-1	-2	-3
	MDD	--	3	2	1	0	0	0	0
	--	MDD	0	0	0	0	1	2	3
None	--	--	0	0	0	0	0	0	0

Step 2: Discrimination and significance.

Based on the BDS of the participants and the disease classification an area- under-the-curve (AUCReal) can be calculated. MedCalc Statistical Software version 16.8 is used for comparison of differences between the ROC curves of groups "serum only", "urine only" and "serum plus urine". The larger the AUCReal, the better the BDS discriminates between healthy and disease and the

more it contains real information for the diagnosis of MDD. The AUCReal is used to determine the optimal pre-set criteria for tail dominance mentioned in Step 1.4 above.

To determine the significance of the discrimination of BDS, 'phenotype randomization' was applied. Phenotype randomization or permutation consists of randomly redistributing the classification of MDD and HC over the original biomarker data: thus the biomarker data per participant are kept unchanged but MDD and HC are permuted at random. Subsequently, a BDSRandom is generated identically as described in Step 1 and the ROC analysis on BDSRandom results in an AUCRandom. Repeating this phenotype randomization 10000 times generates an AUCRandom frequency distribution of which many should be non-discriminative with respect to the disease classification since the relation between biomarker and disease is destroyed by many of the permutations. The p-value is now defined by the fraction of all AUCRandom permutations that are larger or beyond the AUCReal. A p-value below 0.05 indicates a significant discrimination of the BDS and the null hypothesis that the BDS does not discriminate should be rejected.

Step 3. Cross validation and prediction.
Discrimination does not necessarily provide good prediction of new samples. To determine the predictive value of the BDS a (five-fold) cross-validation was performed. The validation was done on the biomarkers that were included by the algorithm on the whole data. The 102 participant cohort was randomly divided into five parts (20, 20, 20, 20 and 22 participants) such that each part contains an equal number of MDDs and HCs. No attempt was made to match age, sex or ethnicity. Five separate sets were constructed: each contains 4 of the 5 parts for training and 1 of the 5 parts for validation and prediction. For each set separately, the participants in the 'training part' were used to determine the cut-offs for the percentiles P1, P5, P10, P90, P95 and P99 as described in Step 1, using the obtained predominance of biomarkers in the full data. Given these cut-offs, a BDS can then be calculated for participants in the validation part together with the ROC-curve for classification of the disease (MedCalc Statistical Software version 16.8 including binomial exact confidence interval for the AUC). To repeat this for each training set we obtain five sets of predicted AUC's (for serum only, urine only, and serum and urine simultaneously). These results provide the predictive value of the BDS to classify MDDs from HCs.

Results

Demographic characteristics

Table 4 shows the demographic characteristics of the subjects that were
included in the 102 participant cohort. Subjects were matched for sex, age and ethnicity. Control subjects had an average HAM-D17 score of 2.7 (range 2-8), while MDD subjects had an average HAM-D17 score of 23.7 (range 11-43; $p<0.0001$).

Due to insufficient amount of serum or identification errors, certain ELISAs were excluded in a minority of participants: 2 biomarkers were tested in all participants, 11 biomarkers in all controls and 50 MDD patients, 5 biomarkers in all controls and 49 MDD patients, and 2 biomarkers in 50 controls and 49 MDD patients.

The results in serum are expressed as a concentration of the biomarker. The results in urine are expressed as the ratio of biomarker to creatinin level. This latter is obtained by dividing the concentration of the biomarker by the concentration of creatinin in urine. As a control for normal renal function, creatinin concentration was measured in serum as well and checked to remain within normal value ranges. Only those within normal serum creatinin range were included.

Table 4. Demographic characteristics.

		All participants	Healthy Controls	MDD	Statistics
Sex	Male	51	22	22	Chi-square
	Female	51	29	29	(2x2): P=0,84
Age (Yr)	Mean	46,6	47,1	46,2	Mann-Whitney-U:
	SD	11,35	11,0	11,5	P=0,81
Ethnicity	Dutch	88	47	41	Chi-square:
	Indonesian	4	2	2	(7x2) P=0,19,
	Surinam	5	2	3	for non-Dutch
	Maroc	2	0	2	(2x2): P=0,15
	Assyric	1	0	1	
	Brazil	1	0	1	
	Iraq	1	0	1	
HAMD17	mean	N.A.	2,7	23,7	p<0,0001
	SD	N.A.	1,1	8,4	

Table 5 shows the Mann-Whitney U and Levene's test for each biomarker tested in serum and in urine. The Mann-Whitney U test, a test for differences in medians, found only a significant difference for Aldosterone in urine and no differences in serum. The Levene's test, a test for differences in variances, found significance for 6 biomarkers, 4 in serum (BDNF, Isoprostane, TNF-R2, Zonulin) and 2 in urine (LTB4 and Thromboxane).

Thus, the traditional (non-parametric) Mann-Whitney U test found only a small number of biomarkers that showed significant differences between MDD and healthy controls and thus limited information would be present to discriminate between these two groups, let alone to predict disease. The test for variability however indicates that there are possible differences in variance that may contribute to the BDS.

BDS and discrimination

The BDS was calculated for 3 groups. Group1 uses only the data of the 21 biomarker levels in serum. Group 2 uses only the data 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. To investigate the pre-set criteria (Step 1.4) for the dominant group with respect to the non-dominant group for exclusion of non-performing biomarker tails, the effect of varying the criteria for P10/P90 and P5/P95 from 0% to 40% was investigated in Group 3 (Serum + Urine).

Putting no requirements on the percentiles, just the dominance classification, the AUC becomes 0.875 without exclusion of any biomarker; Assuming that the mass of the dominant group is at 40% for the P10 and P90, the AUC becomes 0.500 and all biomarkers are excluded. Putting the mass at 20% for the dominant group at P10 and P90 and the mass at 15% at the P5 and P95, the AUC is maximal (AUCReal=0.889). These latter tail criteria exclude 23 biomarkers and include 17 biomarkers (6 in serum (AUCReal=0.755) and 11 in Urine (AUCReal=0.859)) to form the BDS. The included serum biomarkers are

*Table 5. Mann-Whitney U test and Levene's test results of the 2x51 cohort. * Biomarkers with a p<0,05.*

	P-value			
	Serum		Urine	
Biomarkers	Mann-Whitney U test	Levene's test	Mann-Whitney U test	Levene's test
Adiponectin	-	-	0.199	0.435
Aldosterone	0.435	0.501	0.019*	0.720
BDNF	0.595	0.032*	-	-
Calprotectin	0.147	0.170	0.420	0.640
cAMP	0.795	0.137	0.708	0.197
cGMP	-	-	0.799	0.451
Cortisol	0.169	0.094	0.748	0.052
EGF	0.231	0.455	0.392	0.720
Endothelin	0.261	0.098	0.512	0.208
HVEM	0.377	0.759	0.357	0.178
Isoprostane	0.658	0.008*	0.284	0.984
Leptin	0.669	0.223	0.630	0.809
Lipocalin	0.413	0.550	0.239	0.256
LTB4	0.651	0.711	0.928	0.016*
Midkine	0.312	0.992	0.088	0.356
Nitrotyrosin	0.408	0.015	-	-
NPY	0.558	0.782	0.939	0.870
Pregnenolone	-	-	0.846	0.392
Substance P	0.209	0.953	0.139	0.132
Telomerase	0.183	0.301	-	-
Thromboxane B2	0.081	0.666	0.254	0.011*
TNF R2	0.548	0.001*	0.531	0.157
Vitamin D	0.962	0.127	-	-
Zonulin	0.946	0.015*	-	-

TNF-R2, Cortisol, Calprotectin, Thromboxane, Endothelin and Leptin. The included urine biomarkers are cGMP, LN Calprotectin, Leptin, LTB4, Cortisol, Thromboxane, Isoprostane, Aldosterone, HVEM, Midkine and Substance P. Figure 1 visualizes the effect of different exclusion criteria on the AUC and the number of contributing biomarkers in the "serum + urine" group. The pre-set criteria of 20% (for P10/P90) and 15% (for P5/P95) are now fixed throughout further analysis.

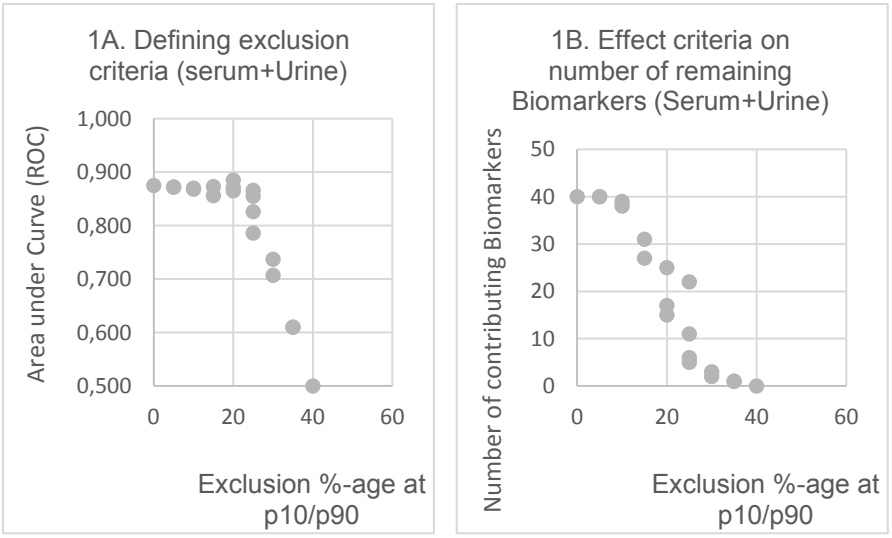
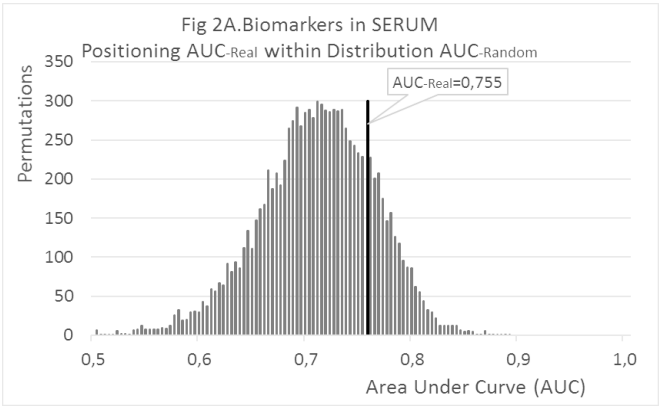
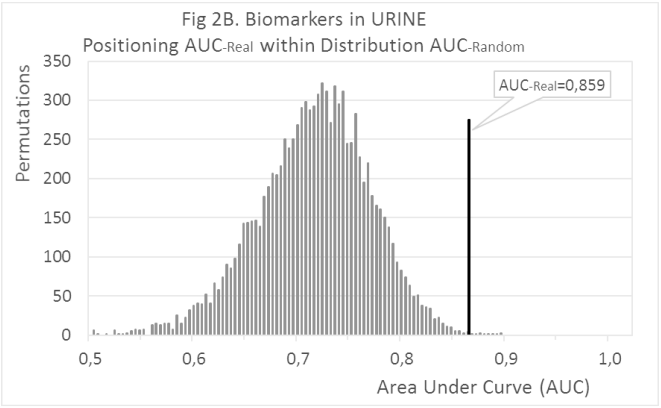


Figure 1. *Effect of exclusion criteria on AUC and number of included biomarkers in the combined group Serum plus Urine. Each point at a p10/p90 exclusion percentage represents the corresponding p5/p95 exclusion condition.*

The phenotype permutation analysis shows a significant BDS discrimination for biomarkers in urine (P<0.001) and for biomarkers in serum and urine together (P<0.001), but not for serum only (P=0.20). The frequency distributions of the AUCRandom, with the AUCReal are visualized for serum, urine, and urine and serum respectively in Figures 2A-C.

The ROC curves are visualized in Figure 3. Comparing ROC curves 'Serum' versus 'Urine': P=0.04; 'Serum' versus 'Serum plus Urine': P=0.0003; 'Urine' versus 'Serum plus Urine' p=0.21.





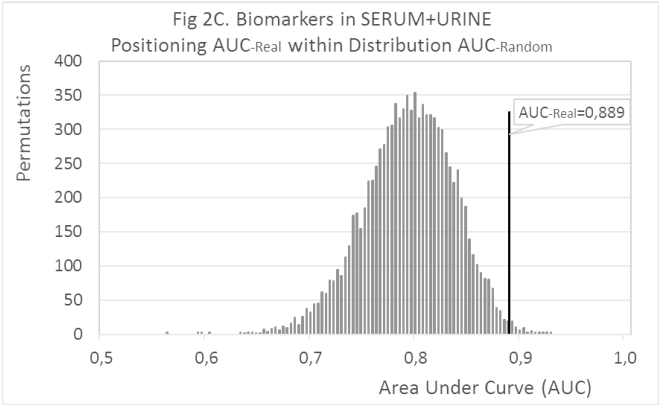


***Figure 2**. Results of phenotype randomization using the exclusion criteria: for P10, P90: 20% and for P5, P95:15%. The position of AUCReal is given. Figure 2A. The 21 Serum biomarkers, histogram of AUCRandom based on 10007 permutations (mean = 0.710; SD = 0.052, fraction right of AUCReal= 19,8%, p-value=0.198); Figure 2B. The 19 Urine biomarkers histogram of AUCRandom based on 10001 permutations (mean = 0.713; SD =0.051, fraction right of AUCReal= 0.02%, P=0.0002); Figure 2C. The combined 40 Serum plus Urine biomarkers histogram of AUCRandom based on 10023 permutations (mean = 0.792; SD = 0.042 fraction right of AUCReal= 0.54%, P=0.005).*

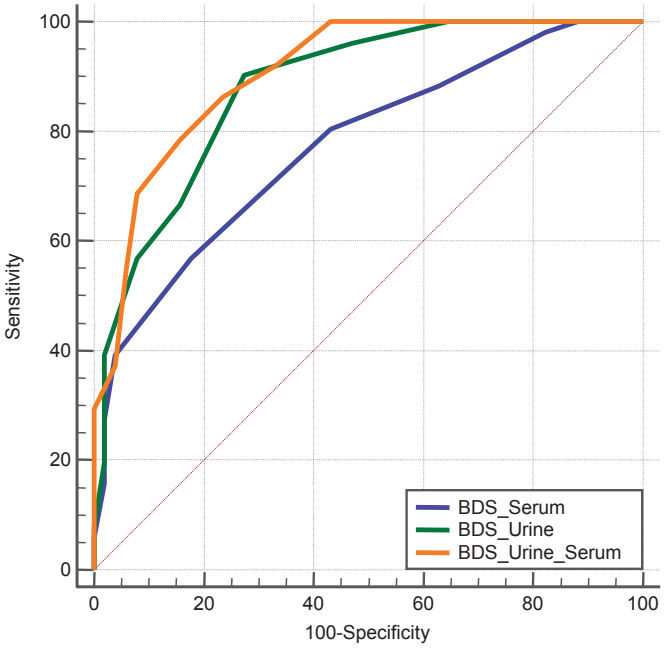


Figure 3. *ROC curves of the Bio Depression Score obtained for serum biomarkers, urine biomarkers, and serum plus urine biomarkers combined (at the pre-set criteria 20% and 15%) using all 102 participants. See text for AUC values.*

Five-fold cross validation

By applying the 20%-15% exclusion criteria, the remaining 6 serum and 11 urine biomarkers, their contributing tails and dominances are fixed. Per training- validation set, the percentile cut-off values in the training subset are determined and applied to the validation subset, leading to predicted AUC-values (Table 6).

The mean AUC of the ROC curves is lowest for the serum Biomarkers, followed by the urine biomarkers and highest in the combined serum plus urine biomarkers. Concommittantly, the confidence intervals in AUC values of ROC curves in the Validation parts range from 0,418 to 0,988 for serum biomarkers, from 0,488 to 0,995 for urine biomarkers, and from 0,569-0,995 for serum plus urine biomarkers. The lowest value is found in serum, followed by Urine and highest in the combination serum plus urine. These predicted results fit with the overall result presented in Figure 1.

Table 6. Results of five-fold cross validaton expressed in the AUC's of the Validation sub sets. V1, V2, V3, V4, V5: the Validation subset in the 1st , 2nd, 3rd, 4th and 5th cross-validation experiment. CI: Confidence interval. BM: biomarker.

	Serum: 6 BMs		**Urine: 11 BMs**		**Serum + urine: 17 BMs**	
	AUC	95% CI	AUC	95% CI	AUC	95% CI
V1	0,660	0,418 - 0,853	0,860	0,633 - 0,972	0,805	0,569 - 0,945
V2	0,905	0,696 - 0,988	0,795	0,558 - 0,940	0,840	0,609 - 0,963
V3	0,907	0,677 - 0,992	0,730	0,488 - 0,901	0,870	0,645 - 0,977
V4	0,731	0,502 - 0,859	0,845	0,615 - 0,966	0,895	0,677 - 0,986
V5	0,777	0,551 - 0,924	0,930	0,736 - 0,995	0,930	0,736 - 0,995
Mean AUC	**0,796**		**0,832**		**0,868**	

Discussion

Diagnosis of MDD may be complicated, especially for non-psychiatrist physicians [4,5]. The aim of this study was to develop a diagnostic biomarker- based test that can aid the physician in diagnosing depression. Based on literature research, early pilot studies and removal of non-contributing biomarkers, we eventually selected 6 serum and 11 urine biomarkers that together can discriminate with high efficiency MDD patients from healthy control subjects, as reflected by an AUC of 0.889 in ROC analysis. Biomarkers in serum appeared less efficient than those in urine. The overall results in serum plus urine were comparable to those in urine only.

To translate the BDS value into clinically applicable results, two routes can be followed. First, a BDS cut-off value can be chosen (to be derived from the AUC- curve), which results in an overall sensitivity / specificity value (for example, from Fig 1. In the serum plus urine AUC-curve at a specificity of 80% a sensitivity of 82% is found). Consequently each individual can be assigned positive or negative. Second, using the AUC characteristic, the BDS value can be translated into a 'chance' of having MDD: values at the horizontal part of the AUC curve have a practically 0% chance of being depressed, whereas values on the vertical part of the curve have a practically 100% chance of being depressed. Values in between have chances ranging gradually from 0-100%.

The advantage of the first approach is simplicity, whereas the advantage of the second approach is the higher precision of diagnosis.

So far, only a handful of studies have combined results of multiple biomarkers into one single diagnostic test [39-44]. A main advantage of the approach used in this study as compared to other depression-related biomarker studies is that we were 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. We show that this method is successful in determining a panel of biomarkers that can separate between MDD patients and controls. The results obtained in this study suggest that this method shows promise for the identification of biomarker panels for other disorders as well.

Furthermore, with the exception of Zheng et al. [44], most studies that have investigated biomarkers for depression focus on serum biomarkers. We however have chosen to investigate biomarker levels in both serum and morning urine in this study, and we have shown an overall superior effect of urine biomarkers compared to serum biomarkers. One of the reasons for this might be that serum levels may show a large variation over the day, limiting their potential clinical use. Biomarker levels in first morning urine may in such cases give more consistent results. The use of urine biomarkers has several advantages over the use of serum biomarkers. 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. In addition, urine is easy to collect and does not require the use of invasive techniques.

Of the previously performed studies investigating combinations of biomarkers for depression, the most consistent results have been obtained by Papakostas *et al*. [42] and Bilello *et al*. [43]. 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% [42]. 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) [43]. This study found a sensitivity of 93%-96% and a specificity of 86%-95%.

Together with these studies, our study demonstrates that the use of biomarker panels is a promising method to increase sensitivity and specificity of biomarker-based tests for depression. Combinations of biomarkers from different "hypotheses" may better represent the heterogenic pathophysiology of depression, and may therefore better reflect disease state [3].

However, several limitations should be kept in mind. 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. Second, although this study indicates that this panel of biomarkers can successfully discriminate between MDD subjects and healthy controls, a second clinically relevant question whether it can also discriminate between MDD and other psychiatric or somatic disorders remains to be answered (MDD specificity). Third, in order to increase cost-effectiveness of this biomarker test to an acceptable level, the number of biomarkers may be reduced, and ideally the final panel should consist of only urine or only serum biomarkers. Fourth, 30 of the 51 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, without taking into account antidepressant use. Despite these limitations, this study is a promising step towards the development of a biomarker-based diagnostic test for depression.

In addition to diagnosis, biomarkers could potentially be used for the identification of those at risk for developing MDD, the prediction of treatment response and the identification of MDD subtypes. The value of our biomarker panel with regard to these applications needs to be determined in future studies. As this panel contains biomarkers related to important pathophysiological mechanisms related to certain MDD subtypes [45-49] it is a promising perspective that MDD subtypes could be identified based on the patient's specific biomarker scores.

Conclusion

We have demonstrated that a panel of biomarkers measured in serum and urine, related to different aspects of MDD pathophysiology, can reliably distinguish MDD subjects from healthy control subjects. Although future research is needed to confirm this result in a separate group of subjects and to optimize the biomarkers included into this panel, this study may be a first step towards the production 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 from such a test.

References

1 Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, et al. Cross- national epidemiology of DSM-IV major depressive episode. BMC Med 2011 Jul 26,;9:90.

2 Sobocki P, Jönsson B, Angst J, Rehnberg C. Cost of depression in Europe. J Ment Health Policy Econ 2006 Jun;9(2):87-98.

3 Jentsch MC, Van Buel EM, Bosker FJ, Gladkevich AV, Klein HC, Oude Voshaar RC, et al. Biomarker approaches in major depressive disorder evaluated in the context of current hypotheses. Biomark Med 2015;9(3):277-297.

4 Cepoiu M, McCusker J, Cole MG, Sewitch M, Belzile E, Ciampi A. Recognition of depression by non-psychiatric physicians--a systematic literature review and meta-analysis. J Gen Intern Med 2008 Jan;23(1):25-36.

5 Su J, Tsai C, Hung T, Chou S. Change in accuracy of recognizing psychiatric disorders by non-psychiatric physicians: five-year data from a psychiatric consultation-liaison service. Psychiatry Clin Neurosci 2011 Dec;65(7):618-623.

6 Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001 Mar;69(3):89-95.

7 Maes M, Galecki P, Chang YS, Berk M. 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 Neuropsychopharmacol Biol Psychiatry 2011 Apr 29,;35(3):676-692.

8 Halaris A. Inflammation, heart disease, and depression. Curr Psychiatry Rep 2013 Oct;15(10):400.

9 Serefko A, Szopa A, Wlaź P, Nowak G, Radziwoń-Zaleska M, Skalski M, et al. Magnesium in depression. Pharmacol Rep 2013;65(3):547-554.

10 Eby GA, Eby KL. Magnesium for treatment-resistant depression: a review and hypothesis. Med Hypotheses 2010 Apr;74(4):649-660.

11 Gold PW, Chrousos GP. Organization of the stress system and its dysregulation in melancholic and atypical depression: high vs low CRH/NE states. Mol Psychiatry 2002;7(3):254-275.

12 Ehlert U, Gaab J, Heinrichs M. Psychoneuroendocrinological contributions to the etiology of depression, posttraumatic stress disorder, and stress-related bodily disorders: the role of the hypothalamus-pituitary-adrenal axis. Biol Psychol 2001 Jul-Aug;57(1-3):141-152.

13 Holsboer F. The corticosteroid receptor hypothesis of depression. Neuropsychopharmacology 2000 Nov;23(5):477-501.

14 Nutt DJ. Relationship of neurotransmitters to the symptoms of major depressive disorder. J Clin Psychiatry 2008;69 Suppl E1:4-7.

15 Ruhé HG, Mason NS, Schene AH. Mood is indirectly related to serotonin, norepinephrine and dopamine levels in humans: a meta-analysis of monoamine depletion studies. Mol Psychiatry 2007 Apr;12(4):331-359.

16 Asberg M, Bertilsson L, Mårtensson B, Scalia-Tomba GP, Thorén P, Träskman-Bendz L. CSF monoamine metabolites in melancholia. Acta Psychiatr Scand 1984 Mar;69(3):201-219.

17 Belmaker RH, Agam G. Major depressive disorder. N Engl J Med 2008 Jan 03,;358(1):55- 68.

18 Cheetham SC, Crompton MR, Czudek C, Horton RW, Katona CL, Reynolds GP. Serotonin concentrations and turnover in brains of depressed suicides. Brain Res 1989 Nov 20,;502(2):332-340.

19 Doris A, Ebmeier K, Shajahan P. Depressive illness. Lancet 1999 Oct 16,;354(9187):1369-1375.

20 Pryor JC, Sulser F. Evolution of the monoamine hypothesis
of depression. In: Horton RW, Katona C, editors. Biological Aspects of Affective Disorders NY, USA: Academic Press; 1991. p. 77-94.

21 Schildkraut JJ. The catecholamine hypothesis of affective disorders: a review of supporting evidence. Am J Psychiatry 1965 Nov;122(5):509-522.

22 Dantzer R, O'Connor JC, Freund GG, Johnson RW, Kelley KW. From inflammation to sickness and depression: when the immune system subjugates the brain. Nat Rev Neurosci 2008 Jan;9(1):46-56.

23 Maes M. A review on the acute phase response in major depression. Rev Neurosci 1993 Oct-Dec;4(4):407-416.

24 Maes M. Evidence for an immune response in major depression: a review and hypothesis. Prog Neuropsychopharmacol Biol Psychiatry 1995 Jan;19(1):11-38.

25 Maes M, Smith R, Scharpe S. The monocyte-T-lymphocyte hypothesis of major depression. Psychoneuroendocrinology 1995;20(2):111-116.

26 Maes M. Major depression and activation of the inflammatory response system. Adv Exp Med Biol 1999;461:25-46.

27 Schiepers OJG, Wichers MC, Maes M. Cytokines and major depression. Prog Neuropsychopharmacol Biol Psychiatry 2005 Feb;29(2):201-217.

28 Videbech P, Ravnkilde B. Hippocampal volume and depression: a meta-analysis of MRI studies. Am J Psychiatry 2004 Nov;161(11):1957-1966.

29 Tanis KQ, Newton SS, Duman RS. Targeting neurotrophic/growth factor expression and signaling for antidepressant drug development. CNS Neurol Disord Drug Targets 2007 Apr;6(2):151-160.

30 Kubera M, Obuchowicz E, Goehler L, Brzeszcz J, Maes M. In animal models, psychosocial stress-induced (neuro)inflammation, apoptosis and reduced neurogenesis are associated to the onset of depression. Prog Neuropsychopharmacol Biol Psychiatry 2011 Apr 29,;35(3):744-759.

31 Deanfield J, Donald A, Ferri C, Giannattasio C, Halcox J, Halligan S, et al. Endothelial function and dysfunction. Part I: Methodological issues for assessment in the different vascular beds: a statement by the Working Group on Endothelin and Endothelial Factors of the European Society of Hypertension. J Hypertens 2005 Jan;23(1):7-17.

32 Kietadisorn R, Juni RP, Moens AL. Tackling endothelial dysfunction by modulating NOS uncoupling: new insights into its pathogenesis and therapeutic possibilities. Am J Physiol Endocrinol Metab 2012 Mar 01,;302(5):481.

33 Eby GA, Eby KL. Rapid recovery from major depression using magnesium treatment. Med Hypotheses 2006;67(2):362-370.

34 Poleszak E. Modulation of antidepressant-like activity of magnesium by serotonergic system. J Neural Transm (Vienna) 2007 Sep;114(9):1129-1134.

35 Poleszak E, Szewczyk B, Kedzierska E, Wlaź P, Pilc A, Nowak G. Antidepressant- and anxiolytic-like activity of magnesium in mice. Pharmacol Biochem Behav 2004 May;78(1):7- 12.

36 Poleszak E, Wlaź P, Kedzierska E, Nieoczym D, Wyska E, Szymura-Oleksiak J, et al. Immobility stress induces depression-like behavior in the forced swim test in mice: effect of magnesium and imipramine. Pharmacol Rep 2006 Sep-Oct;58(5):746-752.

37 Singewald N, Sinner C, Hetzenauer A, Sartori SB, Murck H. Magnesium-deficient diet alters depression- and anxiety-related behavior in mice--influence of desipramine and Hypericum perforatum extract. Neuropharmacology 2004 Dec;47(8):1189-1197.

38 Spasov AA, Iezhitsa IN, Kravchenko MS, Kharitonova MV. Features of central neurotransmission in animals in conditions of dietary magnesium deficiency and after its correction. Neurosci Behav Physiol 2009 Sep;39(7):645-653.

39 Domenici E, Willé DR, Tozzi F, Prokopenko I, Miller S, McKeown A, et al. Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections. PLoS ONE 2010 Feb 11,;5(2):e9166.

40 Arnold SE, Xie SX, Leung Y-, Wang L-, Kling MA, Han X, et al. Plasma biomarkers of depressive symptoms in older adults. Transl Psychiatry 2012 Jan 03,;2:e65.

41 Alawam K, Dudley E, Donev R, Thome J. Protein and peptide profiling as a tool for biomarker discovery in depression. Electrophoresis 2012 Dec;33(24):3830-3834.
42 Papakostas GI, Shelton RC, Kinrys G, Henry ME, Bakow BR, Lipkin SH, et al. Assessment of a multi-assay, serum-based biological diagnostic test for major depressive disorder: a pilot and replication study. Mol Psychiatry 2013 Mar;18(3):332-339.
43 Bilello JA, Thurmond LM, Smith KM, Pi B, Rubin R, Wright SM, et al. MDDScore: confirmation of a blood test to aid in the diagnosis of major depressive disorder. J Clin Psychiatry 2015 Feb;76(2):199.
44 Zheng P, Wang Y, Chen L, Yang D, Meng H, Zhou D, et al. Identification and validation of urinary metabolite biomarkers for major depressive disorder. Mol Cell Proteomics 2013 Jan;12(1):207-214.
45 Lamers F, Vogelzangs N, Merikangas KR, de Jonge P, Beekman ATF, Penninx, B W J H. Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression. Mol Psychiatry 2013 Jun;18(6):692-699.
46 Duivis HE, Vogelzangs N, Kupper N, de Jonge P, Penninx, Brenda W J H. Differential association of somatic and cognitive symptoms of depression and anxiety with inflammation: findings from the Netherlands Study of Depression and Anxiety (NESDA). Psychoneuroendocrinology 2013 Sep;38(9):1573-1585.
47 Becking K, Boschloo L, Vogelzangs N, Haarman BCM, Riemersma-van der Lek R, Penninx, B W J H, et al. The association between immune activation and manic symptoms in patients with a depressive disorder. Transl Psychiatry 2013 Oct 22,;3:e314.
48 Stetler C, Miller GE. Depression and hypothalamic-pituitary-adrenal activation: a quantitative summary of four decades of research. Psychosom Med 2011 Feb-Mar;73(2):114- 126.
49 Vogelzangs N, Duivis HE, Beekman ATF, Kluft C, Neuteboom J, Hoogendijk W, et al. Association of depressive disorders, depression characteristics and antidepressant medication with inflammation. Transl Psychiatry 2012 Feb 21,;2:e79.

[1]Department of Molecular Neurobiology, Center for Life Sciences, University of Groningen, Groningen, The Netherlands,

[2]Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands,

[3]Department of Nuclear Medicine and Molecular Imaging, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands,

[4]Institute of Neuroimmunology and Multiple Sclerosis, Center for Molecular Neurobiology, University Medical Center Eppendorf, Hamburg, Germany