npg

Testing the Predictive Value of Peripheral Gene Expression for Nonremission Following Citalopram Treatment for Major Depression

Jean-Philippe Guilloux[1,2,8], Sabrina Bassi[2,3,8], Ying Ding[4,8], Chris Walsh[2], Gustavo Turecki[5], George Tseng[4,6], Jill M Cyranowski*,[2] and Etienne Sibille*,[2,7]

[1]Université Paris-Sud EA 3544, Faculté de Pharmacie, Châtenay-Malabry, France; [2]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA; [3]Instituto de Ciencias Básicas y Medicina Experimental, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina; [4]Joint CMU-Pitt PhD Program in Computational Biology, University of Pittsburgh, Pittsburgh, PA, USA; [5]McGill Group for Suicide Studies and Depressive Program, Douglas Mental Health Institute, Montréal, QC, Canada; [6]Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, USA; [7]Center for Neuroscience, University of Pittsburgh, Pittsburgh, PA, USA

Major depressive disorder (MDD) in general, and anxious-depression in particular, are characterized by poor rates of remission with first-line treatments, contributing to the chronic illness burden suffered by many patients. Prospective research is needed to identify the biomarkers predicting nonremission prior to treatment initiation. We collected blood samples from a discovery cohort of 34 adult MDD patients with co-occurring anxiety and 33 matched, nondepressed controls at baseline and after 12 weeks (of citalopram plus psychotherapy treatment for the depressed cohort). Samples were processed on gene arrays and group differences in gene expression were investigated. Exploratory analyses suggest that at pretreatment baseline, nonremitting patients differ from controls with gene function and transcription factor analyses potentially related to elevated inflammation and immune activation. In a second phase, we applied an unbiased machine learning prediction model and corrected for model-selection bias. Results show that baseline gene expression predicted nonremission with 79.4% corrected accuracy with a 13-gene model. The same gene-only model predicted nonremission after 8 weeks of citalopram treatment with 76% corrected accuracy in an independent validation cohort of 63 MDD patients treated with citalopram at another institution. Together, these results demonstrate the potential, but also the limitations, of baseline peripheral blood-based gene expression to predict nonremission after citalopram treatment. These results not only support their use in future prediction tools but also suggest that increased accuracy may be obtained with the inclusion of additional predictors (eg, genetics and clinical scales).

Neuropsychopharmacology (2015) **40**, 701–710; doi:10.1038/npp.2014.226; published online 1 October 2014

INTRODUCTION

Major depressive disorder (MDD) is a common psychiatric disorder, affecting an estimated 19 million people each year in the United States alone (Kessler *et al*, 2003, 2005; Murray and Lopez, 1997). For patients with MDD, the lifetime mortality due to suicide approaches 15% (Mann, 2003), and the annual cost associated with MDD-related disability and death ranks in the tens of billions of dollars in the US (Kupfer *et al*, 2012; Murray and Lopez, 1997). A critical barrier to

devise optimal treatment and achieve sustained remission is the lack of predictive tests for disease state and treatment remission. The uncomplicated, nonintrusive, and cost-effective availability of blood samples make blood-based biomarkers leading candidates for large-scale development.

The potential of blood transcriptome (the panel of detectable gene transcripts) for monitoring health and disease states is being investigated for peripheral organ and brain diseases (for review, see (Mohr and Liew, 2007)). Convergent functional genomic approaches have proposed biomarkers for mood states in bipolar depression (Le Niculescu *et al*, 2008). Padmos *et al* (2008) have reported a profile of proinflammatory gene transcripts that discriminated bipolar-depressed patients from controls. Pajer *et al* (2012) differentiated between MDD and control subjects, and between MDD with and without anxiety in adolescents, using candidate blood and brain gene expression from animal models of depressive-like behaviors. Using inflammatory, stress and neuroplasticity markers, Papakostas *et al* (2011) developed a serum-based protein assay that robustly

*Correspondence: Dr E Sibille, Departments of Psychiatry and Pharmacology, University of Toronto, Campbell Family Research Institute, Toronto, ON, Canada M5T 1R8, Tel: +416 535 8501 ext 33542, Fax: +416 979 4704, E-mail: etienne.sibille@camh.ca or Dr JM Cyranowski, Chatham University, 1 Woodland Road, Pittsburgh, PA 15232, USA, Tel: +412 365 1568, Fax: +412 365 1130, E-mail: j.cyranowski@chatham.edu
[8]These authors contributed equally to the work.

discriminated MDD from control patients, and that replicated in an independent validation cohort. Similarly, leukocyte expression of proinflammatory, neuroendocrine stress axis and neuroplasticity markers discriminated between MDD and control subjects; and baseline proinflammatory genes were differentially expressed between responders and nonresponders to two antidepressant treatments (Nortryptyline or escitalopram) (Cattaneo *et al*, 2012). Using a microarray approach, Mamdani *et al* (2011) found that changes in the expression of 32 probesets predicted changes in Hamilton Rating Scale for Depression (HRSD) scores following citalopram treatment. Belzeaux *et al* (2012) reported baseline changes between responder and nonresponder MDD subjects and identified four gene transcripts with predictive value for treatment response.

In this study, we tested the potential of gene expression to predict nonremission prior to treatment initiation. We focused on nonremission, as it significantly contributes to the chronic illness burden suffered by many patients. Specifically, we hypothesize that MDD patients who do not achieve symptom remission after 12 weeks of treatment (MDD-NR, MDD-nonremitter) differ in baseline gene expression from the subjects who will remit (MDD-R, MDD-remitter). If so, baseline blood gene expression should predict remission after treatment with citalopram, a common antidepressant. We measured large-scale blood transcriptome changes in 34 MDD and 33 healthy control subjects (initial cohort), as part of an ongoing study of anxious-depression treated with a 12-week course of combined citalopram and weekly psychotherapy treatment. In the first phase, we performed an exploratory analysis of differential expression to seek evidence for group differences, namely MDD-NR and MDD-R *vs* controls, being cognizant of the high false positive rate associated with these approaches. In the second and independent phase, we used the same dataset to construct a predictive model with a machine learning method with support vector machines and corrected the results for the bias of reporting the 'best' prediction rates. We then sought to confirm the prediction results using the existing clinical and blood gene expression data in a second cohort of MDD subjects with similar, although not identical characteristics, and who were treated for 8 weeks with citalopram alone (validation cohort) (Mamdani *et al*, 2011). We also had two sets of secondary predictions; first, we predicted that baseline gene expression profiles of MDD nonremitters would be characterized by an upregulation of genes associated with immune activation and inflammation. Second, we predicted that the inclusion of clinical scales would improve prediction of nonremission.

MATERIALS AND METHODS

A short summary is provided here. Details and technical information on the cohorts and analyses are in the online supplements. The initial cohort included 34 anxious-depressed adults recruited for an outpatient trial of combined medication and psychotherapy treatment, and an age- and gender-matched sample of 33 nondepressed controls (Table 1). The patient sample met criteria for a nonpsychotic MDD episode of sufficient severity (documented by DSM-IV SCID interview and score ⩾15

on the 25-item HRSD (Hamilton, 1960)) and elevated symptoms of panic or anxiety (score ⩾7 on the past-month panic and agoraphobic spectrum selfreport (Cassano *et al*, 1997; Shear *et al*, 2001)). Demographic and clinical scales are listed in Supplementary Table S1. Nonpatient controls did not meet criteria for any mood or anxiety disorder. Individuals with a history of schizophrenia, bipolar disorder or antisocial personality disorder; drug or alcohol abuse/ dependence in the past 3 months; severe, uncontrolled medical illness; history of nonresponse to citalopram; and women who were pregnant or planning to become pregnant were excluded from participation. Subjects were free of antidepressant treatment prior to the start of the study. All study procedures were approved by the University of Pittsburgh Institutional Review Board. Treatment for the MDD cohort included citalopram (average dose at week 12, 37.5 ± 10.5 mg) and weekly psychotherapy with either (randomly assigned) interpersonal psychotherapy for depression with panic and anxiety symptoms (IPT-PS) or brief supportive psychotherapy (BSP) (Hellerstein *et al*, 1998). Analyses were conducted on an intent-to-treat basis, with depression remission defined as achieving three consecutive weeks of average HRSD-17 scores ⩽7. The trial is registered at ClinicalTrials.gov under NCT00930293.

Blood was collected in PAXgene tubes at study entry (T0) and after 12 weeks of treatment (T12) and processed onto Illumina HT12-v4.0 gene array. Gene functional analysis was performed using Ingenuity Pathway Analysis (http:// www.ingenuity.com). Analysis of promoter transcription factor enrichment was performed with Enrichr (Chen *et al*, 2013). Real-time quantitative polymerase chain reaction (qPCR) was performed on a Mastercycler real-time PCR machine (Eppendorf, Hamburg, Germany) using universal PCR conditions (Sibille *et al*, 2009). Differential expression was analyzed by random intercept model with variable selection using Bayesian information criteria and class label permutation (Wang *et al*, 2012). Prediction analyses for nonremission were performed using standard machine learning method with support vector machines (with linear kernel) protocol. To correct for the selection bias of reporting the best model, we added a nested cross-validation step (Tibshirani, 2009; Varma and Simon, 2006).

For the validation cohort, we used clinical and transcriptome data from a published study (Table 1; Mamdani *et al* (2011) for details), in which MDD patients were treated with citalopram (10–60 mg, titration based on response and tolerability) for 8 weeks, with depression remission defined as HRSD-17 score ⩽7 at week 8 assessment. Blood samples were collected at T0 and T8 in PAXgene tubes and processed onto U133 + 2.0 Affymetrix GeneChips. The top genes identified in the initial cohort predictions were used to construct the prediction model for response/remission in the validation cohort.

RESULTS

Table 1 provides the baseline demographic and clinical features of the two cohorts used in the report. A total of 34 MDD and 33 nondepressed control subjects comprised the initial cohort. Six MDD subjects dropped out of treatment prior to the week 12 (T12) blood draw. Only one of these six

**Table 1** Clinical and Demographic Factors of the Initial Cohort

Combined cohort	Test cohort									
	T0					T12				
	MDD	**CTRL**	**Statistic**	*P*-value	df	**MDD**	**CTRL**	**Statistic**	*P*-value	df
Age	30.27 (*10.31*)	34.01 (*11.15*)	t = − 1.42	0.16	64.2					
Sex	27 (F)/7 (M)	27 (F)/6 (M)	$\chi^2=0$	1.00	1					
Race	28^W/3^B/2^BW/1^PI	28^W/5^B	$\chi^2=3.49$	0.32	3					
Marital	22^A/6^B/6^D	20^A/9^B/2^C/2^D	$\chi^2=4.68$	0.20	3					
Education	8^α/13^β/13^γ	7^α/5^β/21^γ	$\chi^2=5.49$	0.06	4					
HRSD-17	18.47 (*2.88*)	1.36 (*1.60*)	t = 30.15	1.58E-34	51.8	8.61 (*4.81*)	2.44 (*3.73*)	t = 5.54	1.09E-06	50.3
HRS-A	16.33 (*4.65*)	1.13 (*1.36*)	t = 17.97	4.33E-20	37.7	7.75 (*4.37*)	1.91 (*2.45*)	t = 6.30	1.68E-07	40.6
QIDS-16	14.29 (*3.70*)	1.54 (*1.20*)	t = 19.06	1.08E-21	40.1					
MDD-R cohort	**MDD-R**	**CTRL (R)**	**Statistic**	*P*-value	df	**MDD-R**	**CTRL (R)**	**Statistic**	*P*-value	df
Age	28.66 (*9.64*)	31.72 (*9.92*)	t = − 0.94	0.36	34.0					
Sex	13 (F)/5 (M)	13 (F)/5 (M)	$\chi^2=0$	1	1					
Race	15^W /2^BW/1^PI	16^W/2^B	$\chi^2=5.03$	0.17	3					
Marital	11^A/5^B/2^D	14^A/4^B/0^D	$\chi^2=2.47$	0.29	2					
Education	4^α/8^β/6^γ	4^α/3^β/11^γ	$\chi^2=3.74$	0.15	4					
HRSD-17	17.50 (*2.81*)	1.11 (*1.08*)	t = 23.08	7.39E-17	21.9	5.83 (*2.31*)	2.67 (*2.91*)	t = 3.62	0.001	32.3
HRS-A	15.76 (*4.38*)	1.06 (*1.23*)	t = 13.29	5.77E-11	18.7	5.50 (*2.68*)	2.00 (*1.49*)	t = 4.83	4.94E-05	26.6
QIDS-16	11.94 (*3.00*)	1.56 (*1.20*)	t = 13.64	2.64E-12	22.3					
MDD-NR cohort	**MDD-NR**	**CTRL (NR)**	**Statistic**	*P*-value	df	**MDD-NR**	**CTRL (NR)**	**Statistic**	*P*-value	
Age	32.08 (*11.12*)	35.87 (*12.21*)	t = − 0.86	0.40	25.8					
Sex	13 (F)/1 (M)	13 (F)/1 (M)	$\chi^2=0$	1	1					
Race	11^W/3^B	11^W/3^B	$\chi^2=0$	1	1					
Marital	9^A/1^B/4^D	6^A/4^B/2^C/2^D	$\chi^2=5.07$	0.17	3					
Education	4^α/4^B/6^γ	3^α/2^B/9^γ	$\chi^2=1.41$	0.49	3					
HRSD-17	19.64 (*2.71*)	1.71 (*2.13*)	t = 19.49	1.77E-16	24.6	13.60 (*4.03*)	1.50 (*2.91*)	t = 7.69	7.92E-07	16.4
HRS-A	16.93 (*5.00*)	1.07 (*1.49*)	t = 11.37	7.27E-09	15.3	11.80 (*3.91*)	1.50 (*3.06*)	t = 6.56	4.85E-06	17.0
QIDS-16	17.29 (*2.40*)	1.43 (*1.22*)	t = 22.03	3.73E-15	19.3					

	Validation cohort	
	T0	**T8**
Age	38.68 (*11.65*)	
Sex	35 (F)/28 (M)	
Education	12.4^δ	
HRSD-17	30.06 (*6.47*)	10.35 (*7.28*)

Abbreviations: HRS-A, Hamilton Rating Scale for Anxiety; HRSD-17, 17-item Hamilton Rating Scale for Depression; QIDS16, Quick Inventory of Depressive Symptomatology.
Values are mean (± SD). Group codes are as following Sex: F, female; M, male; Self-declared race: ^W white, ^B black, ^BW black,white, ^PI pacific islander; Marital status: ^A never married; ^B married; ^C living together as married; ^D separated/divorces. Education: ^α <high school, high school, GED, technical school; ^β some college; ^γ college; ^δ average years of education.

met remission criteria prior to drop out, and was thus classified as a remitter; the remaining five were classified as nonremitters in intent-to-treat prediction analyses, in line with CONSORT clinical trial reporting standards (http://www.consort-statement.org/). Of those who remained at T12, 18 were classified as remitters (MDD-R) and 10 as nonremitters (MDD-NR). Thus, 19 of 34 subjects (55.9%) were classified as treatment remitters in intent-to-treat analyses. At baseline assessment, MDD-R and MDD-NR patients did not significantly differ in age, race, education,
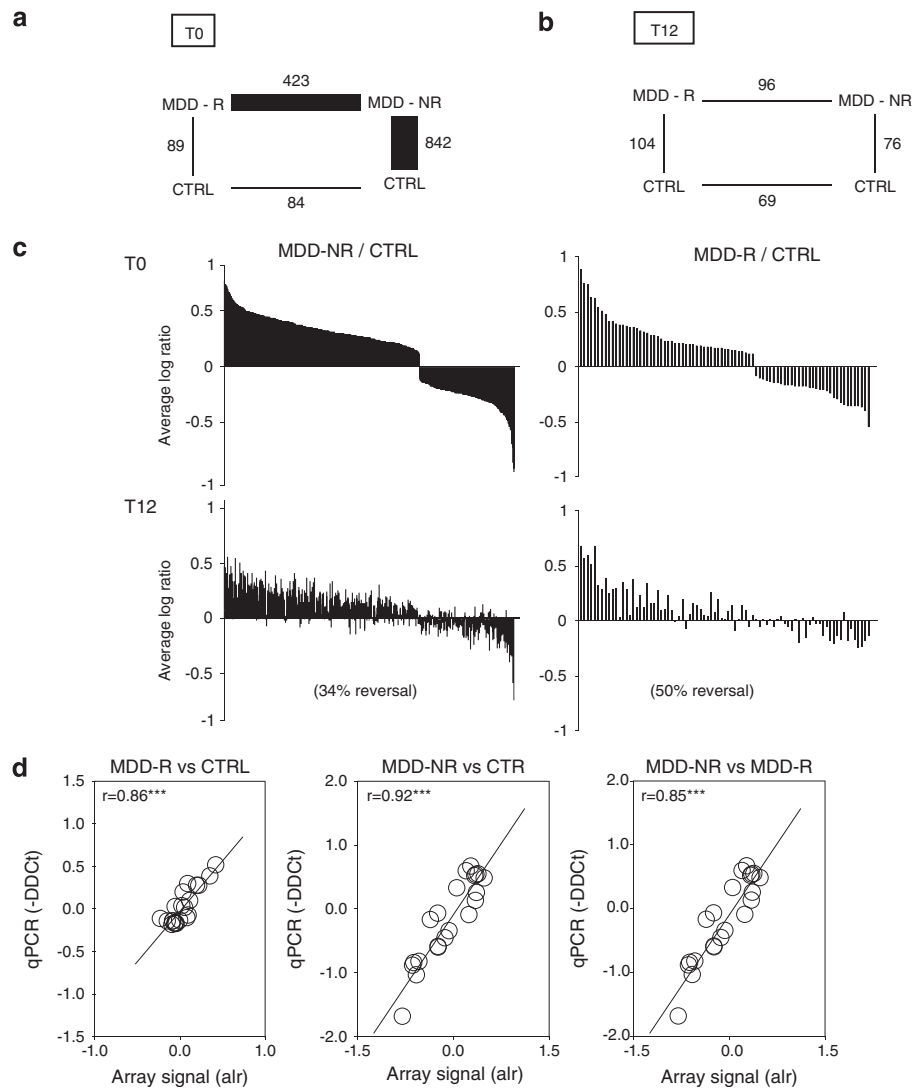
Figure 1 Peripheral gene expression in MDD. (a) At T0, MDD subjects who will not remit (MDD-NR) displayed a greater number of differentially expressed genes compared with controls (CTRL) and to MDD subjects who will remit (MDD-R). The line thickness indicates weighted numbers of differentially expressed genes between groups. The two control groups (CTRL) correspond to subjects matched to the respective MDD groups (Table 1). (b) At T12, the number of differentially expressed genes was reduced between MDD-NR and other groups, and was at the level observed between control groups, suggesting that it reflected background gene variability. (c) The T0 profiles of changes in gene transcript levels for MDD-related genes are reversed by antidepressant treatments. T0, upper panels: MDD-related genes are organized by the extent of their changes in expression along the x axis ($p < 0.01$, FC > 20%). Vertical bars indicate the extent of changes [average $\log_2$ of MDD-NR/CTRL (Left) or MDD-R/CTRL (Right) expression ratios]. T12, lower panels: vertical bars indicate the relative transcript levels for the same MDD-related genes after citalopram/psychotherapy treatments. Genes are sorted in the same order as in upper panels. The T12 profiles correspond to a 34% reversal in MDD-NR and to a 50% reversal in MDD-R subjects (both $p < 0.0001$). (d) Validation of array results by independent qPCR measurements in MDD-R and MDD-NR subjects. See Supplementary Tables S7-8 for details. Alr: average log ratio.

anxiety (HRSA scores), or gender distribution (all *p*-values > 0.10). However, the MDD-NR group did display higher baseline clinician-rated and self-reported depression scores as compared with the MDD-R group (HRSD-17 = 19.6 *vs* 17.5, *p* = 0.02; Quick Inventory of Depressive Symptom scores, QIDS = 17.3 *vs* 11.9, *p* < 0.01; Table 1).

Blood Gene Expression of MDD-NR Subjects Differ from MDD-R and Control Subjects at Baseline

In the initial cohort, results from exploratory analyses of differential gene expression at baseline (pretreatment; T0)

between MDD and control subjects are summarized in Figure 1a. Few changes were observed between MDD-R and matched control subjects, compared with between MDD-NR and matched control subjects. Gene changes in MDD-R subjects did not correlate with effects in MDD-NR subjects (Pearson correlation factor, $R = -0.05$), and a direct comparison between MDD-R and MDD-NR revealed intermediate number of differentially expressed genes. False positives and/or normal background variability in gene expression was assessed as differences observed between subgroups of healthy (nondepressed) control subjects. Results from these exploratory analyses suggest that at T0,

Table 2 Biological Functions Over-represented in Differentially-expressed Gene Groups

Group comparisons	Numbers of biological groups at q < 0.01	Biological groups
(TO) MDD-NR/CTRL (Details in Supplementary Table S3)	154	Neutrophils: function, homing, chemotaxis, cell movement. Myeloid cells: response, cell movement, phagocytosis. Granulocytes: homing, chemotaxis, cell movement. Leukocytes: function, cell movement, migration, chemotaxis, quantity, proliferation. Phagocytes: phagocytosis, immune response, function, engulfment, homing, cell movement, response of Lymphocytes: proliferation. Immune Cells: proliferation, adhesion. Cells: chemotaxis, homing, immune response, phagocytosis Viral Infection, function of blood cells. Apoptosis of lymphoma cell lines, Inflammatory response.
(TO) MDD-NR/MDD-R (Details in Supplementary Table S4)	12	Viral Infection, replication of virus, replication of RNA virus, function of leukocytes, cell rolling of phagocytes, replication of Influenza A virus, function of neutrophils, cell death of myeloid cells, apoptosis of myeloid cells, infection of mammalia, immune response of cells, invasion of cells.
(TO) MDD-R/CTRL	0	None
(TO) CTRL(NR)/CTRL(R)	0	None
(T12) MDD-NR/CTRL	0	None
(T12) MDD-NR/MDD-R	0	None
(T12) MDD-R/CTRL	0	None
(T12) CTRL(NR)/CTRL(R)	0	None

Significant biological groups defined using an ingenuity pathway analysis were regrouped and summarized by cell types. See Supplementary Table S4 for details.

(a) untreated MDD subjects differ from control subjects, (b) MDD differences are mostly driven by MDD-NR subjects, and (c) MDD-NR subjects differ from MDD-R subjects (Figure 1a). See Supplementary Table S2 for details on individual genes.

An analysis of cell-specific markers between MDD-NR and control subjects did not suggest that these findings were supported by changes in cell-type densities (Supplementary Results and Supplementary Table S3). Biological functions identified as over-represented in MDD-NR patients at T0 compared with controls corresponded almost exclusively to increased immune function, inflammation, and white blood cell recruitment (Table 2; details in Supplementary Table S4). Moreover, transcription factor binding sites enriched in promoter regions of differentially expressed genes were similarly linked to immune and inflammatory responses, including NF-kB-related signaling and cell cycle regulation, and/or mediation of tumor growth/suppression (Supplementary Table S6).

qPCR was performed for 21 genes with T0 differences (ADSL, ARHGEF11, CD3D, ECGF1 (TYMP), GBA, GZMA, IFITM3, IL17RA, MBOAT1, MPST, MSH6, NAP1L4, OSCAR, PRDM4, RPL4, RPL5, RPL17, RPL24, RSPH3, TBXAS1, and TIMM23). The qPCR results correlated with array data for the MDD-R vs control (Pearson correlation $r = 0.86$, $p = 5e^{-7}$), MDD-NR vs control ($r = 0.92$, $p < 1e^{-7}$) and MDD-NR vs MDD-R ($r = 0.85$, $p = 5e^{-7}$) (Figure 1d), together supporting the technical reliability of the array data, although not all individual assays reached significance. Results from individual genes are in Supplementary Tables S7-8. Results from post-treatment (T12) blood gene expression and treatment effects are summarized in Figure 1b. All T12 groups differences were at the level of background gene variability, as measured between control subjects matching MDD-R or MDD-NR subjects. Comparing T0 and T12, most genes reverted back

toward control expression levels after treatment (ie, closer to x axis in Figure 1c).

Together, the results of those exploratory approaches suggest that nonremitter subjects are biologically different from remitter subjects, hence providing a rationale for performing an independent prediction analysis for non-remission.

Prediction of Nonremission After Treatment

To predict depression remission at T12 based on T0 transcriptome data, we constructed a predictive model with a machine learning method with support vector machines. As standard practice for machine learning protocols, we started with the T0 dataset, ranked all genes based on statistical significance, and filtered the data based on effect sizes in the training set. We then systematically constructed prediction models using the top sets of genes, ranging from 2–30 genes, and applied these models to the test set (Figure 2). This process is iterated until each sample is treated as test set once. Typically, the prediction accuracy is reported for an optimal gene set; however, to address the common bias that is introduced by reporting the 'best result', we added an additional step to correct the results for model selection, using a nested cross-validation approach (Tibshirani, 2009; Varma and Simon, 2006) (Figure 2). The results from these studies indicate an average cross-validated accuracy (ie, model selection bias corrected) of 79.4% in predicting remission status, with the 13-gene model displaying the highest individual noncorrected prediction value (88%) (Tables 3 and 4 and Supplementary Table S9). An analysis of correlation of expression with cell-specific markers suggests that the 13-gene set is potentially, although not exclusively, enriched in genes expressed in lymphocytes (Supplementary Results and Supplementary
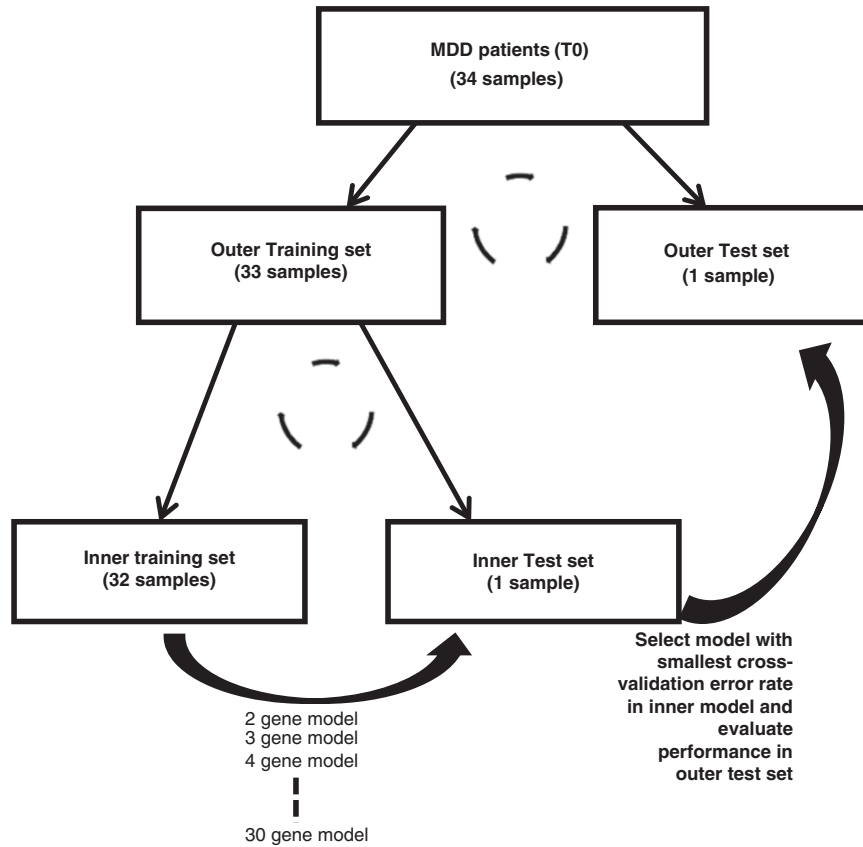
Figure 2 Flowchart of blood biomarker prediction performance assessment, including correction of model selection bias. To search for the optimal number of features (2–30 genes) in the prediction model, we applied a nested cross-validation (CV) with two nested CV loops. The dataset is initially divided into two: one sample as the test set, and the second as the training set. Then a leave-one-out cross validation (LOOCV) method is applied on the training set using all classifiers (top 2–30 genes). The classifiers with the smallest error rates are selected and used to build the model within the training set. The model is then evaluated on the left-out test sample. This procedure is repeated until all samples have been left out once. This means that the left-out test sample is independent from the model selection stage, including the selection of the model with the minimum error rate. Together, this procedure guarantees an unbiased error estimate.

Table 3 Prediction of Nonremission in MDD Patients at Baseline

Features used in prediction model	Initial cohort (genes only)	Validation cohort (genes only)
# features used in model	13	6
Corrected accuracy (%)	79.4	76.2
Sensitivity (%)	66.7	86.1
Specificity (%)	89.5	59.3

Accuracy represents the percentage of correctly identified samples. Sensitivity represents the ability to predict nonremission. Specificity represents the ability to predict remission. The corrected accuracy was obtained by nested cross-validation approach as described in Figure 2.

Table S10), highlighting the value of using whole blood transcriptome for predictive purposes.

Validation of Gene-based Prediction of Nonremission in an Independent Cohort

Given the small size and unique features of the initial cohort and the potential bias in model selection, we sought to

confirm the prediction model in an independent cohort recruited and tested at another institution. We obtained an archival dataset drawn from a study that evaluated gene expression in blood from 63 MDD subjects who were subsequently treated for 8 weeks with similar doses of citalopram (see Table 1 and Mamdani *et al* (2011) for details). Thus, this study represented a close, but not identical, match to our initial cohort. We built a new prediction model in the validation cohort using the same 13 genes identified in the initial cohort, and found through another round of leave-one-out cross validation that a 6-gene model achieved the highest accuracy (76.2%) in the validation cohort (Table 4 and Supplementary Table S11).

Exploratory Analyses

Predictive models including clinical data. We explored the potential of clinical features to augment the gene-based prediction models. These analyses were performed in the initial cohort only, because of the availability of multiple clinical scales in that study (Supplementary Table S1). We first applied machine learning prediction methods using clinical scale data only, considering models with 2–30

Table 4 Top Genes Included in the Prediction Models

Gene symbol	Gene name	Frequency of use in initial cohort	Frequency of use in validation icohort
CD3D	CD3d molecule, delta (CD3-TCR complex)	34	46
CD97	CD97 molecule	34	
IFITM3	Interferon induced transmembrane protein 3	34	63
RPL5	Ribosomal protein L5	34	
GZMA	Granzyme A	33	63
TAGLN2	Transgelin 2	33	
TIMP1	TIMP metallopeptidase inhibitor 1	33	
RPL24	Tibosomal protein L24	29	63
PSMA4	Proteasome (macropain) subunit, alpha type, 4	25	
MATR3	Matrin 3	18	63
RPL9	Ribosomal protein L9	15	
PSMA6	Proteasome (macropain) subunit, alpha type, 6	12	
RPL17	Ribosomal protein L17	7	58

Genes in bold were included in the independent qPCR validation (Figure 1). Consistent with the previous functional and transcription factor analyses, this 13-gene model included genes associated with immune/inflammatory activation (CD3D, CD97, IFITM3, and GZMA) and mediation of cell proliferation (GZMA and TIMP1), although not exclusively.

clinical variables. The following three baseline clinical variables provided optimal accuracy in predicting T12 remission status, with corrected accuracy of 70.6% as determined by nested cross validation (as in Figure 2): clinician-rated depressive symptoms (HRSD-17 scores), patient-reported depressive symptoms (QIDS (Rush *et al*, 2003)), and perceived stress scores (4-item Perceived Stress Scale (Cohen *et al*, 1983)). Next we tested models including both clinical and gene expression data and show that T12 depression nonremission could be predicted with a corrected accuracy of 97% based on T0 gene expression and clinical data (Supplementary Tables S12-13), in a model including just three features: two genes (IFITM3 and TIMP1) and one clinical feature (QIDS). Notably the type of adjunctive psychotherapy treatment provided in that cohort in addition to citalopram (IPT-PS or BSP) did not contribute to any of the prediction models.

Treatment response trajectory. Because MDD-NR and MDD-R patients differed significantly in baseline depression scores (Table 1), we explored whether observed differences in posttreatment remission status represented differences in treatment trajectory. Figure 3 shows a large overlap in baseline depression scores between the MDD-NR and MDD-R groups, and significant group differences in slopes between T0 and T12 in the initial cohort ($p < 0.00137$) and T0 and T8 in the validation cohort ($p < 1.25e^{-6}$), hence confirming the presence of a different trajectory of treatment response in MDD-NR subjects. Consistent with the overlap in values (Figure 3), baseline HRSD-17 values provided moderate prediction for nonremission (67.6% in initial and 57.1% in validation cohort).

DISCUSSION

Results from the first and exploratory phase of this study suggest that at pretreatment assessment, the gene expression profiles obtained from blood samples of MDD subjects who will not attain remission after treatment (MDD-NR group) differ from nondepressed controls and also from MDD patients who will remit with treatment (MDD-R) (Figure 1). The nature of genes affected in MDD-NR subjects suggests a state of inflammation and immune system activation in those patients, although this was not confirmed by independent approaches (Table 2). In the second phase of the study, using a machine learning prediction model and a leave-one-out cross validation to correct for multiple model testing, we show that pretreatment baseline gene expression predicted nonremission after treatment with 79.4% accuracy (Table 3), based on a 13-gene panel (Table 4). A strength of this study is that we confirmed our results in an independent validation cohort. Specifically, we showed that 6 out of 13 genes identified in the initial cohort could predict remission in an independent cohort ($n = 63$ MDD patients) with 76.2% accuracy (Table 3). Together, this report demonstrates the potential, but also the limitations, of pretreatment peripheral gene expression profiles to predict nonremission following an 8- to 12-week course of citalopram treatment.

Although the sample size of the initial cohort was small, the gene expression prediction results replicated in an independent cohort, despite differences in experimental design (8- *vs* 12-week drug exposure, no psychotherapy) and inclusion of nonanxious MDD patients in the validation cohort. It is worth noting that we did not apply the optimized 13-gene model from our cohort directly to the validation cohort study, but instead used those top 13 candidate markers to perform a new model selection in the independent cohort with feature selection and leave-one-out cross validation to assess the prediction accuracy. This approach was designed to facilitate the analysis and mitigate differences across the two cohorts and platforms, and it may explain why genes differentially expressed in specific cohorts (see introduction) may not necessarily be included in the final common set of predictors. For future
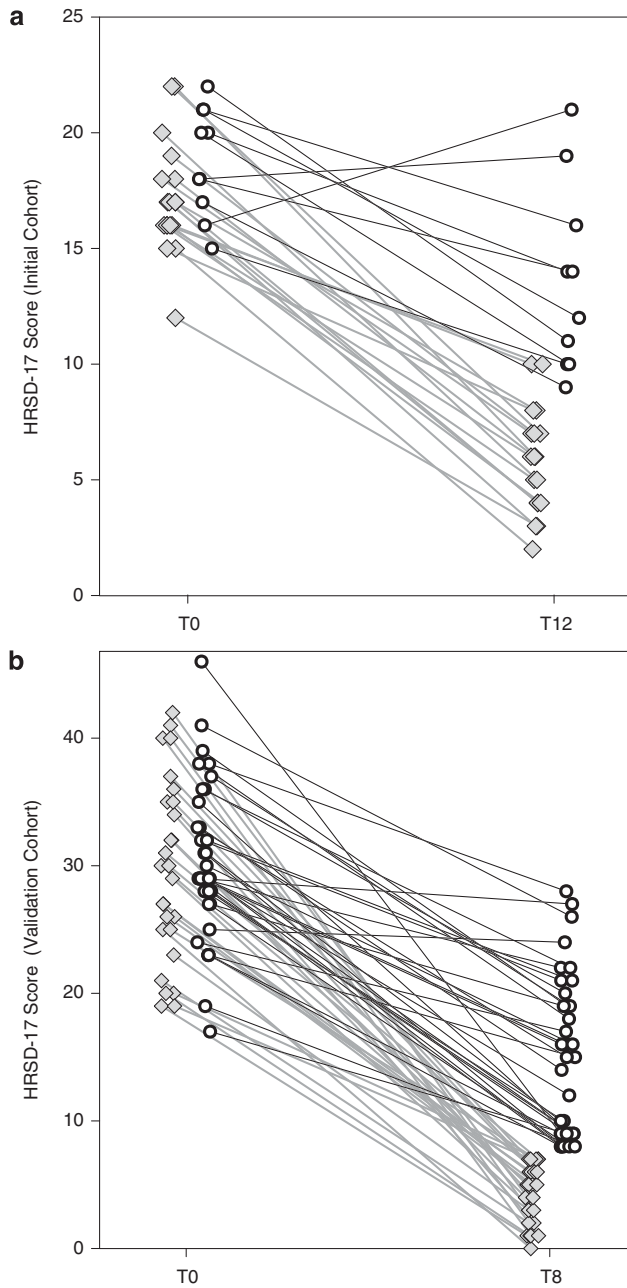
Figure 3 Trajectories of HRSD17 scores between MDD-NR and MDD-R subjects in the initial and validation cohorts. Grey squares indicate MDD-R subjects. Empty black circles indicate MDD-NR subjects.

translational and clinical utility, a randomized prospective clinical trial using a preestablished predictive model is the ultimate validation. To date, however, direct application of predictive models to different test studies has been difficult due to cross-laboratory variability in probe designs from different array platforms and different experimental protocols. Such issues could be addressed in the future through the use of carefully standardized experimental protocols or commercialized products. For example, prognostic prediction of breast cancer treatment (Cheng et al, 2009; Shen et al, 2004; Xu et al, 2008) has been greatly facilitated by the use of MammaPrint (van't Veer et al, 2002) and Oncotype

DX (Paik et al, 2006), two successful microarray and qPCR examples with cross-laboratory standardization.

The combination of clinical scales and gene expression data enhanced the accuracy of the nonremission prediction to 97% in the initial cohort (Table 3 and Supplementary Table S11). Notably, of all clinical scales included in the combined machine learning models (including those assessing current and lifetime mood and anxiety symptoms, childhood trauma, and social/interpersonal function; see supplements and Supplementary Table S1), the only scale retained in the combined prediction model was the QIDS-16; a patient-reported indicator of depressive symptom severity. This finding may highlight the importance of assessing subjective levels of patient-reported distress which may, moreover, be more closely linked to over-activation of pro-inflammatory pathways (Cole, 2010; Miller et al, 2008; Pace and Miller, 2009).

What could we learn from the exploratory analyses of differential gene expression? We are cognizant of the high false positive rate associated with the uncorrected statistical approaches used here, so we briefly discuss results at the gene group level rather than for independent genes. As hypothesized, biological functions represented by genes showing baseline expression differences largely suggest pre-treatment elevations in inflammation and immune activation in MDD-NR subjects (Table 2). This may, in part, relate to baseline elevations in symptoms of depression and of perceived stress in the MDD-NR group. In this regard, MDD-NR subjects also displayed heightened expression of numerous transcripts with response elements for NF-kB, a proinflammatory transcription factor associated with the experience of chronic stress (Cole, 2010; Miller et al, 2008; Pace and Miller, 2009). The fact that the prevalence of these predictive gene markers was reduced after treatment despite elevated HRSD scores (Table 2) suggests that while inflammation and immune-related pathways could have contributed to MDD onset, their presence may not be necessary for the maintenance of MDD in those subjects. Alternatively, putative remodeling of brain circuitry and other molecular/neurochemical changes (such as changes in glucocorticoid receptor sensitivity) may sustain the depressive state despite the absence of continued peripheral immune and inflammatory gene markers. Finally, it is worth noting that the extent of correlation of expression between the predictive genes and cell-specific markers suggests the contribution of various cell types. Future studies may address the contribution of changes within and across specific leukocyte cell populations, but the current results suggest that using total blood gene transcripts may be appropriate for predictive purpose.

The finding that MDD-NR subjects showed differential gene expression and trajectories of reduction in HRSD-17 scores following treatment (Figure 3) supports the intrinsic difference of this group compared with other MDD subjects. Future studies will need to characterize those differences on longer timescales and to examine the association between the observed pretreatment gene expression profiles and functional indicators of systemic inflammation or stress-induced immune/inflammatory reactivity (Maes, 1995; Zunszain et al, 2013). Conversely, subjects who remitted after treatment (MDD-R) had pretreatment gene expression profiles similar to background variability in control

subjects, demonstrating that diagnostic MDD status is not necessarily accompanied by peripheral gene changes (at least with the technical sensitivity of the gene arrays used here). This observation highlights the underlying biological heterogeneity in MDD subjects, and underscores the need for future mental health biomarker research to move beyond traditional diagnostic categories, as articulated in the NIMH Research Domain Criteria initiative (Insel *et al*, 2010).

Together, the current and prior studies (see introduction) not only support the biomarker potential of peripheral gene expression in predicting MDD treatment outcome but also highlight the need for additional studies. Future studies will need to be performed in larger cohorts and may need to combine measures of gene expression with selected clinical scales, specifically related to mood symptoms and chronic life stress. The inclusion of genetic information, measures of peripheral metabolites and treatment history may further increase the predictive value of those assays. For instance, Gudayol-Ferre *et al* (2013) showed that a combination of clinical, genetic (val/Met COMT polymorphism) and neuro-psychological variables were associated with a pattern of remission in MDD subjects. The fact that in our study baseline peripheral gene expression did not differ between MDD subjects who remitted and healthy controls also suggests that different sets of biomarkers may be needed to capture the heterogeneity of MDD subjects. The findings from this study show the potential of blood gene expression to identify prior to treatment—with high sensitivity and specificity—a population of individuals that would benefit from treatment augmentation. Future studies will need to assess means of treatment augmentation in the context of prediction of nonremission. Whether gene expression biomarkers will provide insight into disease mechanisms for a rational choice of drug augmentation is currently unknown, so the added value of those assays will have to be determined based on optimized design and better knowledge of successful augmenting factors, but the levels of accuracy and confirmation obtained under the current conditions are promising. Finally, future treatment paradigms will benefit from determining whether other antidepressants are successful in treating the nonremitting patients identified in this study.

Notes and Limitations

Several limitations to the current study are noteworthy. First, the initial cohort sample size was small and thus potentially associated with high false positive rates for both differential expression and prediction analyses. Second, with the exception of the Mamdani *et al* (2011) study, direct comparisons of prediction results with prior reports (see Introduction) were not performed due to differences in study goals (eg, disease status *vs* outcome prediction here). Third, the lack of available anxiety measures in the validation cohort may have limited the analysis in that cohort. Fourth, we did not attempt to separate blood cells and instead investigated transcriptome profiles at the time of collection, so that expression of putative markers would not be influenced by sample manipulation. Instead, using prior results, we ruled out the possible interference of blood cell proportions on differential expression of genes. Fourth, the

biological validity of the inflammation and immune activation was not confirmed by other biological tests and is thus speculative at this time. Fifth, the study was not designed to assess later time points regarding delayed remission and potential biomarkers for future relapse.

DISCLAIMER

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIMH or the National Institutes of Health.

REFERENCES

Belzeaux R, Bergon A, Jeanjean V, Loriod B, Formisano-Treziny C, Verrier L *et al* (2012). Responder and nonresponder patients exhibit different peripheral transcriptional signatures during major depressive episode. *Transl Psychiatry* **2**: e185.

Cassano GB, Michelini S, Shear MK, Coli E, Maser JD, Frank E (1997). The panic-agoraphobic spectrum: a descriptive approach to the assessment and treatment of subtle symptoms. *Am J Psychiatry* **154**(6 Suppl): 27–38.

Cattaneo A, Gennarelli M, Uher R, Breen G, Farmer A, Aitchison KJ *et al* (2012). Candidate genes expression profile associated with antidepressants response in the GENDEP study: differentiating between baseline 'predictors' and longitudinal 'targets'. *Neuropsychopharmacology* **38**: 377–385.

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV *et al* (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform* **14**: 128.

Cheng C, Shen K, Song C, Luo J, Tseng GC (2009). Ratio adjustment and calibration scheme for gene-wise normalization to enhance microarray inter-study prediction. *Bioinformatics* **25**: 1655–1661.

Cohen S, Kamarck T, Mermelstein R (1983). A global measure of perceived stress. *J Health Soc Behav* **24**: 385–396.

Cole SW (2010). Elevating the perspective on human stress genomics. *Psychoneuroendocrinology* **35**: 955–962.

Gudayol-Ferre E, Guardia-Olmos J, Pero-Cebollero M, Herrera-Guzman I, Camarena B, Cortes-Penagos C *et al* (2013). Prediction of the time-course pattern of remission in depression by using clinical, neuropsychological, and genetic variables. *J Affect Disord* **150**: 1082–1090.

Hamilton M (1960). A rating scale for depression. *J Neurol Neurosurg Psychiatry* **23**: 56–62.

Hellerstein DJ, Rosenthal RN, Pinsker H, Samstag LW, Muran JC, Winston A (1998). A randomized prospective study comparing

supportive and dynamic therapies. Outcome and alliance. *J Psychother Pract Res* **7**: 261–271.

Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K et al (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* **167**: 748–751.

Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR et al (2003). The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *J Am Med Assoc* **289**: 3095–3105.

Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* **62**: 593–602.

Kupfer DJ, Frank E, Phillips ML (2012). Major depressive disorder: new clinical, neurobiological, and treatment perspectives. *Lancet* **379**: 1045–1055.

Le Niculescu H, Kurian SM, Yehyawi N, Dike C, Patel SD, Edenberg HJ et al (2008). Identifying blood biomarkers for mood disorders using convergent functional genomics. *Mol Psychiatry* **14**: 156–174.

Maes M (1995). Evidence for an immune response in major depression: a review and hypothesis. *Prog Neuropsychopharmacol Biol Psychiatry* **19**: 11–38.

Mamdani F, Berlim MT, Beaulieu MM, Labbe A, Merette C, Turecki G (2011). Gene expression biomarkers of response to citalopram treatment in major depressive disorder. *Transl Psychiatry* **1**: e13.

Mann JJ (2003). Neurobiology of suicidal behaviour. *Nat Rev Neurosci* **4**: 819–828.

Miller GE, Chen E, Sze J, Marin T, Arevalo JM, Doll R et al (2008). A functional genomic fingerprint of chronic stress in humans: blunted glucocorticoid and increased NF-kappaB signaling. *Biol Psychiatry* **64**: 266–272.

Mohr S, Liew CC (2007). The peripheral-blood transcriptome: new insights into disease and risk assessment. *Trends Mol Med* **13**: 422–432.

Murray CJ, Lopez AD (1997). Global mortality, disability, and the contribution of risk factors: global burden of disease study. *Lancet* **349**: 1436–1442.

Pace TW, Miller AH (2009). Cytokines and glucocorticoid receptor signaling. Relevance to major depression. *Ann N Y Acad Sci* **1179**: 86–105.

Padmos RC, Hillegers MH, Knijff EM, Vonk R, Bouvy A, Staal FJ et al (2008). A discriminating messenger RNA signature for bipolar disorder formed by an aberrant expression of inflammatory genes in monocytes. *Arch Gen Psychiatry* **65**: 395–407.

Paik S, Tang G, Shak S, Kim C, Baker J, Kim W et al (2006). Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* **24**: 3726–3734.

Pajer K, Andrus BM, Gardner W, Lourie A, Strange B, Campo J et al (2012). Discovery of blood transcriptomic markers for depression in animal models and pilot validation in subjects with early-onset major depression. *Transl Psychiatry* **2**: e101.

Papakostas GI, Shelton RC, Kinrys G, Henry ME, Bakow BR, Lipkin SH et al (2011). Assessment of a multi-assay, serum-based biological diagnostic test for major depressive disorder: a pilot and replication study. *Mol Psychiatry* **18**: 332–339.

Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN et al (2003). The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* **54**: 573–583.

Shear MK, Frank E, Rucci P, Fagiolini DA, Grochocinski VJ, Houck P et al (2001). Panic-agoraphobic spectrum: reliability and validity of assessment instruments. *J Psychiatr Res* **35**: 59–66.

Shen R, Ghosh D, Chinnaiyan AM (2004). Prognostic meta-signature of breast cancer developed by two-stage mixture modeling of microarray data. *BMC Genom* **5**: 94.

Sibille E, Wang Y, Joeyen-Waldorf J, Gaiteri C, Surget A, Oh S et al (2009). A molecular signature of depression in the amygdala. *Am J Psychiatry* **166**: 1011–1024.

Tibshirani R (2009). A bias correction for the minimum error rate in cross-validation. *Ann Appl Stat* **3**: 822–829.

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M et al (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**: 530–536.

Varma S, Simon R (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinform* **7**: 91.

Wang X, Kang DD, Shen K, Song C, Lu S, Chang LC et al (2012). An R package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection. *Bioinformatics* **28**: 2534–2536.

Xu L, Tan AC, Winslow RL, Geman D (2008). Merging microarray data from separate breast cancer studies provides a robust prognostic test. *BMC Bioinform* **9**: 125.

Zunszain PA, Hepgul N, Pariante CM (2013). Inflammation and depression. *Curr Top Behav Neurosci* **14**: 135–151.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (http://www.nature.com/npp)