Archival Report

Biological Psychiatry

Replicable and Coupled Changes in Innate and Adaptive Immune Gene Expression in Two Case-Control Studies of Blood Microarrays in Major Depressive Disorder

Gwenaël G.R. Leday, Petra E. Vértes, Sylvia Richardson, Jonathan R. Greene, Tim Regan, Shahid Khan, Robbie Henderson, Tom C. Freeman, Carmine M. Pariante, Neil A. Harrison, MRC Immunopsychiatry Consortium, V. Hugh Perry, Wayne C. Drevets, Gayle M. Wittenberg, and Edward T. Bullmore

ABSTRACT

BACKGROUND: Peripheral inflammation is often associated with major depressive disorder (MDD), and immunological biomarkers of depression remain a focus of investigation.

METHODS: We used microarray data on whole blood from two independent case-control studies of MDD: the GlaxoSmithKline–High-Throughput Disease-specific target Identification Program [GSK-HiTDiP] study (113 patients and 57 healthy control subjects) and the Janssen–Brain Resource Company study (94 patients and 100 control subjects). Genome-wide differential gene expression analysis (18,863 probes) resulted in a p value for each gene in each study. A Bayesian method identified the largest p-value threshold (q = .025) associated with twice the number of genes differentially expressed in both studies compared with the number of coincidental case-control differences expected by chance.

RESULTS: A total of 165 genes were differentially expressed in both studies with concordant direction of fold change. The 90 genes overexpressed (or UP genes) in MDD were significantly enriched for immune response to infection, were concentrated in a module of the gene coexpression network associated with innate immunity, and included clusters of genes with correlated expression in monocytes, monocyte-derived dendritic cells, and neutrophils. In contrast, the 75 genes underexpressed (or DOWN genes) in MDD were associated with the adaptive immune response and included clusters of genes with correlated expression in T cells, natural killer cells, and erythroblasts. Consistently, the MDD patients with overexpression of UP genes also had underexpression of DOWN genes (correlation $>$ .70 in both studies).

CONCLUSIONS: MDD was replicably associated with proinflammatory activation of the peripheral innate immune system, coupled with relative inactivation of the adaptive immune system, indicating the potential of transcriptional biomarkers for immunological stratification of patients with depression.

Keywords: Affymetrix, Bayesian, Biomarker, Inflammation, Systems, Transcriptome

http://dx.doi.org/10.1016/j.biopsych.2017.01.021

Depression and inflammation are often associated with one another. Depressive symptoms in a large population sample were significantly related to blood concentrations of C-reactive protein (CRP; odds ratio ~1.8 for depressive symptoms in people with CRP $>$ 3 mg/L vs. CRP $<$ 1 mg/L) (1). Multiple case-control studies of major depressive disorder (MDD) have reported increased peripheral blood concentrations of CRP (Cohen's d ~0.50) and proinflammatory cytokines such as interleukin 6 (d ~0.50) and tumor necrosis factor (d ~0.40) in MDD (2,3). The prevalence of comorbid depression is increased in many nonpsychiatric inflammatory disorders (4).

There is growing evidence for a causal effect of inflammation on depression. Peripheral inflammation precedes the emergence of depressive symptoms in longitudinal epidemiological studies (5) and in about 30% of patients receiving proinflammatory interferon-α treatment for hepatitis C (6,7). Experimental challenge with peripheral proinflammatory stimuli in animals robustly induces a syndrome of illness behavior and anhedonia that approximates depressive symptoms (8). Peripheral immune cells and cytokines are known to mediate signals across the blood-brain barrier by several mechanisms (9). Activation of microglia can locally amplify the effects of even a weak peripheral proinflammatory signal on neuronal function and behavior (10).

These observations suggest that pharmacological disruption of peripheral proinflammatory signals could be

ISSN: 0006-3223

therapeutically effective, at least for a subgroup of patients with depression. However, it is most unlikely that any single anti-inflammatory drug will prove to be superior to all existing treatments for all patients (11). Only about a third of patients with depression have biological evidence for peripheral inflammation [e.g., CRP > 3 mg/L (12)], and an anti-inflammatory drug seems likely a priori to be most effective for an inflamed subgroup of patients with depression. There are many markers of the peripheral immune system that can be conveniently measured in venous blood samples from patients with depression, including cytokines, CRP, and other proteins; cell counts from flow cytometry; and gene transcription (13).

Transcriptional (messenger RNA [mRNA]) biomarkers have the potential advantages of assay stability (compared with cytokines) and target specificity (compared with cell counts). However, previous case-control studies of peripheral blood gene expression and MDD have been inconsistent (14–18) (Supplemental Table S1). To date, the Netherlands Study of Depression and Anxiety (NESDA) (14) is the largest single case-control study of the peripheral blood transcriptome in MDD (882 current patients with MDD, 635 remitted patients with MDD, and 331 control subjects). The NESDA reported statistically significant (false discovery rate [FDR] = 10%) differential expression of 129 genes enriched for interleukin 6 and natural killer (NK) cell signaling pathways (19).

We were primarily motivated by the hypothesis that MDD is associated with peripheral blood transcriptional markers of innate immune system activation (8,20). We were also concerned to focus on results that were more likely to replicate across case-control studies of gene expression in MDD. We report Affymetrix microarray data on 18,863 probes from two independently designed and conducted case-control studies of MDD. We used a Bayesian method to identify genes that were differentially expressed in both studies. Focusing on a consensus set of 165 genes, we investigated the functional significance of the genes that were differentially (over- or under-) expressed in cases compared with controls. We also explored the secondary hypothesis that innate immune system activation is coupled to relative inactivation of the adaptive immune system in patients with MDD (21–23).

METHODS AND MATERIALS

Samples

We analyzed data from two case-control studies of depression: the GlaxoSmithKline–High-Throughput Disease-specific target Identification Program (GSK–HiTDiP) study and the Janssen–Brain Resource Company (Janssen–BRC) study. Other aspects of these studies have been previously reported (24–28); demographic and clinical details on the samples are provided in Supplemental Table S2.

GSK–HiTDiP. This study was designed primarily to test for an association between genetic (DNA) variation and diagnosis of depression. Minimal sociodemographic and clinical data were collected, but microarray data were available for analysis (after quality control) from whole-blood samples stored for less than 6 years on a sample comprising 113 patients with MDD prospectively balanced for comorbid anxiety disorder [57 with generalized anxiety disorder diagnosed by the Mini-International Neuropsychiatric Interview (29) and 56 without anxiety disorder] and 57 healthy control subjects. All participants provided informed consent in writing. The study was approved by an independent ethics review board.

Janssen–BRC. This study was designed primarily for biomarker discovery. Microarray data were available for analysis (after quality control) from whole-blood samples stored for less than 1 year on a sample comprising 94 patients with MDD (40 with generalized anxiety disorder diagnosed post hoc by the Mini-International Neuropsychiatric Interview and 54 without anxiety disorder) and 100 healthy control subjects. Additional data on melancholic symptom severity, anxiety, substance use, and body mass index (BMI) were available for patients with MDD. All participants provided informed consent in writing. The study was approved by an independent ethics review board.

Whole-blood samples from both studies were analyzed using the Affymetrix Human Genome U133 plus 2.0 array. We applied identical quality control, normalizing, and annotation algorithms to the two datasets, resulting in the estimation of mRNA expression at each of 18,863 unique probes for each participant.

Differential Expression Analysis

To determine differential gene expression between cases and controls, we adopted the same strategy for both studies; see Supplemental Figure S1 for a schematic overview of the data analysis strategy. For each gene $i = 1, \ldots, 18{,}863$, we fitted a linear regression model that included group (coded Gr; two-level factor, case/control), batch (B; two-level factor, 1/2), gender (Ge; two-level factor, male/female), age (Ag; continuous), and presence/absence of anxiety (An; two-level factor, 0/1) as covariates. Denoting samples by $j = 1, \ldots, n$ and the design matrix by X, the model is

$$y_{ij} = \beta_{i0} + \sum_{k=1}^{2} \beta_{ik}^{Gr} x_{jk}^{Gr} + \sum_{l=1}^{2} \beta_{il}^{B} x_{jl}^{B} + \sum_{m=1}^{2} \beta_{im}^{Ge} x_{jm}^{Ge} + \beta_{i}^{Ag} x_{j}^{Ag} + \sum_{r=1}^{2} \beta_{ir}^{An} x_{jr}^{An} + \varepsilon_{i}, \quad (1)$$

where $\varepsilon_i \sim N(0, \sigma_i^2)$. For identifiability, we imposed the contrasts $\beta_{i1}^{Gr} = \beta_{i1}^{B} = \beta_{i1}^{Ge} = \beta_{i1}^{An} = 0$ on model parameters. The model was fitted using the R package limma (30). Subsequently, we tested the null hypothesis H_{i0}: $\beta_{i2}{}^{Gr} = 0$; that is, there is no difference in expression of the ith gene between the two groups using the moderated t statistic (31). For each of the 18,863 probes in each study, the p value was generated using the asymptotic approximate distribution (and also nonparametrically by a permutation algorithm; see Supplemental Tables S3 and S4).

Combining *p* Values for Differential Gene Expression From Two Studies

To identify MDD-related genes that were replicated in both the GSK–HiTDiP and Janssen–BRC datasets, we set the p-value threshold for significance of differential expression of each gene in each study to optimize in some sense the number of