ORIGINAL ARTICLE

Type I interferon signaling genes in recurrent major depression: increased expression detected by whole-blood RNA sequencing

S Mostafavi[1], A Battle[1], X Zhu[2], JB Potash[3], MM Weissman[4], J Shi[5], K Beckman[6], C Haudenschild[7], C McCormick[8], R Mei[9], MJ Gameroff[4], H Gindes[3], P Adams[4], FS Goes[10], FM Mondimore[10], DF MacKinnon[10], L Notes[11], B Schweizer[10], D Furman[12], SB Montgomery[13,14], AE Urban[2], D Koller[1] and DF Levinson[2]

A study of genome-wide gene expression in major depressive disorder (MDD) was undertaken in a large population-based sample to determine whether altered expression levels of genes and pathways could provide insights into biological mechanisms that are relevant to this disorder. Gene expression studies have the potential to detect changes that may be because of differences in common or rare genomic sequence variation, environmental factors or their interaction. We recruited a European ancestry sample of 463 individuals with recurrent MDD and 459 controls, obtained self-report and semi-structured interview data about psychiatric and medical history and other environmental variables, sequenced RNA from whole blood and genotyped a genome-wide panel of common single-nucleotide polymorphisms. We used analytical methods to identify MDD-related genes and pathways using all of these sources of information. In analyses of association between MDD and expression levels of 13 857 single autosomal genes, accounting for multiple technical, physiological and environmental covariates, a significant excess of low P-values was observed, but there was no significant single-gene association after genome-wide correction. Pathway-based analyses of expression data detected significant association of MDD with increased expression of genes in the interferon α/β signaling pathway. This finding could not be explained by potentially confounding diseases and medications (including antidepressants) or by computationally estimated proportions of white blood cell types. Although cause–effect relationships cannot be determined from these data, the results support the hypothesis that altered immune signaling has a role in the pathogenesis, manifestation, and/or the persistence and progression of MDD.

INTRODUCTION

Major depressive disorder (MDD) is a complex phenotype with moderate heritability estimates (31–42%).[1] Physiological correlates include alterations in neurohormonal secretion, structural brain anatomy, and markers of immune function and inflammation. In contrast to the other two most-studied psychiatric disorders in adults (bipolar disorder and schizophrenia), or traits with comparable heritability such as Type 2 diabetes (26%),[2] it has been challenging to identify relevant genetic factors for MDD from genome-wide association studies (GWAS), with the largest GWAS to date (with over 9000 MDD cases) detecting no significant associations with common single-nucleotide polymorphisms (SNPs).[3] This might be explained by MDD's phenotypic heterogeneity[4,5] and complex genetic architecture,[6] and its association with environmental factors that are likely to interact with genetics.[7]

Whole-transcriptome studies offer another type of genome-wide search for disease-related mechanisms by measuring mRNA expression levels of each gene in a relevant tissue. Although expression data do not directly disentangle cause–effect relationships, altered expression levels in disease can reflect the effect of common and/or rare genetic sequence variation, environmental factors, the effects of the disease processes itself, and interaction between genetic variation and environmental factors. We applied deep RNA sequencing (RNA-seq) to whole-blood RNA in a large sample from a population-based survey research panel, including cases on and off psychiatric medication (unlike most clinical samples). This ascertainment method, which made whole blood the only feasible tissue for expression profiling, had the advantage of providing information on subjects' "real-life" state (in contrast with cell lines or postmortem tissues), with the limitations of requiring statistical correction of many state-related,

[1]Department of Computer Science, Stanford University, Stanford, CA, USA; [2]Department of Psychiatry and Behavioral Science, Stanford University School of Medicine, Stanford, CA, USA; [3]Department of Psychiatry, University of Iowa Hospitals & Clinics, Iowa City, IA, USA; [4]Department of Psychiatry, Columbia University and New York State Psychiatric Institute, New York, NY, USA; [5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA; [6]Biomedical Genomics Center, University of Minnesota, Minneapolis, MN, USA; [7]Personalis, Inc., Menlo Park, CA, USA; [8]Illumina, Inc., La Jolla, CA, USA; [9]Centrillion Biosciences, Inc., Palo Alto, CA, USA; [10]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD, USA; [11]Depatment of Clinical Psychology, American University, Washington DC, DC, USA; [12]Department of Microbiology & Immunology, School of Medicine, Stanford University, Stanford, CA, USA; [13]Department of Genetics, Stanford University, Stanford, CA, USA and [14]Department of Pathology, Stanford University, Stanford, CA, USA. Correspondence: Professor Dr DF Levinson, Department of Psychiatry and Behavioral Science, Stanford University School of Medicine, 401 Quarry Rd, Room 3322, 94305-5797, Stanford, CA, USA.
E-mail: dflev@stanford.edu
Received 24 May 2013; revised 27 August 2013; accepted 24 September 2013; published online 3 December 2013

possibly confounding variables, and of having limited potential to identify some brain-specific mechanisms. On the other hand, there is increasing evidence implicating dysregulation of glucocorticoid and immune responses in MDD, and white blood cells may be a particularly a suitable tissue for studying the relevant immunological factors.

We present here the largest whole-transcriptome study of MDD to date and the first using RNA-seq. After collecting psychiatric, demographic, environmental and medical information, we studied 922 European ancestry individuals (463 cases and 459 controls) with RNA-seq of whole-blood RNA and with a GWAS assay. Three types of analysis of association to MDD were then carried out: (i) analyses of individual gene expression levels produced a significant excess of low *P*-values, but no significant association of a single gene after correction for genome-wide testing; (ii) analyses of expression levels across pre-defined pathways and gene sets detected significant genome-wide association with the interferon (IFN) α/β signaling pathway; and (iii) joint evaluation of gene expression and SNP genotypes identified significant association to *CINP* (cyclin-dependent kinase 2 interacting protein), a gene involved in the cell cycle arrest, possibly induced by IFN.[8,9]

MATERIALS AND METHODS

Subject recruitment and assessment
(See Supplementary Methods and Batte *et al.*,[10] (in revision, Genome Research) for additional methodological details.) Recruitment was carried out under Institutional Review Board-approved protocols. Cases (recurrent or chronic MDD) and controls were recruited by Knowledge Networks (Menlo Park, CA, USA) from a nationally representative panel of ~60 000 individuals. Knowledge Networks invited 14 463 individuals (self-reported Caucasian ancestry, ages 21–60) for screening. Respondents were screened online with lifetime self-report versions of the depression and alcohol/substance dependence modules of the Composite International Diagnostic Interview-Short Form (CIDI-SF)[11] plus screening items for psychotic and bipolar disorders. After excluding those who reported current dependence or lifetime psychotic or bipolar disorder, individuals who tentatively met inclusion criteria (see below) were invited to participate. Those who consented to be contacted answered additional online questions about height, weight, childhood trauma and smoking. Among those who completed blood draw by a national phlebotomy company (after giving written informed consent), 650 prospective cases and 589 prospective controls were interviewed by telephone with a modified Structured Clinical Interview for DSM-IV (full depression, bipolar, alcohol, substance and anxiety modules; psychosis screen; and a more detailed illness and medication history); Patient Health Questionnaire (PHQ-9) (current depression)[12] and Generalized Anxiety Disorder 7-item scale (GAD-7) (current anxiety)[13] scales; and a family history screen (MDD, bipolar disorder and suicide). Most telephone interviews were completed within 1–2 months of online screening and blood draw, with delays of several months in a small proportion of cases. (Sample sizes at each stage are provided in Supplementary Methods.)

Clinical data were reviewed by the clinical site principal investigator (PI) (MMW or JBP) and the overall PI (DFL). After exclusions for substance dependence, non-affective psychosis and bipolar disorder, we identified 475 cases with MDD with two or more lifetime episodes or one episode lasting ⩾2 years, and 474 controls who never experienced a 2-week period with depressed mood plus two or more other MDD criteria. After exclusions for genotypically non-European ancestry, unusual medical comorbidities, and RNA-seq and GWAS quality control analyses, 922 individuals were included in further analyses (463 cases and 459 controls).

RNA-seq and genotype data
Whole blood was collected in PaxGene tubes (Qiagen, Valencia, CA, USA) for RNA and in acid-citrate-dextrose tubes for DNA. PaxGene tubes contain a proprietary cell lytic agent for immediate release of RNA from cells; mRNA levels remain stable for years with appropriate storage.[14] Extracted DNA was genotyped with the Illumina Omni1-Quad microarray (Centrillion Biosciences, Mountain View, CA, USA). Post-quality control genotypes were available for 720 591 autosomal SNPs. RNA was extracted; most globin RNA was removed (GLOBINclear Kits, Life Technologies, Grand Island, NY, USA)

(Supplementary Figures S1–S4); and Illumina TrueSeq kits (Illumina, Inc., La Jolla, CA, USA) were used for RNA purification (polyA selection), chemical fragmentation, single-stranded cDNA conversion, DNA library preparation and oligonucleotide barcoding. Sequencing was carried out with an Illumina HiSeq 2000 (Illumina, Inc.; 50- or 51-bp single-ended reads, 3 multiplexed libraries per lane) yielding 70 million reads (average) per individual. Reads were mapped to the NCBI v37 *Homo sapiens* reference genome using Tophat[15] Complete genotype and RNA-seq data are available form https://nimhgenetics.org (see data access note below).

Gene expression was quantified with HTSeq using the 'intersection_strict' criteria.[16] Only uniquely aligned reads were used to quantify gene expression levels. Reads were assigned to 21 578 of the 22 339 annotated protein-coding genes (NCBI v37). Analyses included 13 857 autosomal genes with ⩾100 individuals with ⩾10 reads (in total from anywhere across the transcript). Effects of technical covariates (for example, per individual 5′ bias, GC bias, sequencing depth, and percent globin reads) and biological covariates (for example, estimates of blood cell-type proportions and time of day of blood draw; see Supplementary Table S1 for complete list) were removed by ridge regression of logarithm-transformed read counts. Cell-type proportions were inferred using a method based on non-negative least squares,[17] making use of external microarray data on cell-type specific expression signatures[18] (Supplementary Methods).

Association of gene expression levels with MDD
A likelihood ratio test (LRT) was used to determine the significance of association between the expression levels of each gene and MDD. Using LRT, the strength of the association is determined by comparing the likelihood of the null ('background') model that includes only a set of confounding factors (see below) with the likelihood of the full model (confounding factors plus expression value; Supplementary Methods). Final *P*-values were obtained using permutation analyses (8000 initial permutations or 1 000 000 permutations for three genes with $P = 0$ in the initial 8000) (Supplementary Figure S5). False discovery rate (FDR) was used for multiple hypothesis correction.[19]

In total, we corrected for 39 'background' covariates (Table 1, Supplementary Table S2). These consist of 24 environmental, physiological and medical covariates, including body mass index, gender, smoking and age; manually curated indicators of intake of medications and substances (reported by ⩾30 individuals), including cholesterol-lowering and antihypertensive medication classes, thyroid medication, cannabis and alcohol use, many of which were more common in cases and had a non-negligible effect on gene expression (Supplementary Table S2 and Supplementary Figure S10). Other covariates include five genotypic principal components (PCs) reflecting population structure[20] (Supplementary Methods and Supplementary Figure S4); and 10 hidden confounding factors obtained as PCs of the residual expression data (Supplementary Figure S6). We decided to include expression PCs because previous studies have shown that removal of expression PCs results in improved power for detecting expression quantitative trait locus (eQTLs)[21] and identifying true-positive co-expression relationships between genes.[22] We also evaluated whether there was an excess of low *P*-values (deviation from the expected uniform distribution), by estimating the proportion of true positives using the π_1 statistic.[23]

Association of expression of gene pathways
Association of gene pathways was analyzed using two methods, based on all 1325 canonical pathways from MSigDB[24] (c2.cp.v3.1) that contained 5–100 expressed genes in our data. Gene set enrichment analysis (GSEA)[25] assessed the significance of each pathway using 5000 permutations of MDD status for each pathway, where the log(LRT *P*-values) was used as the score for each gene. Hypergeometric tests were used to evaluate over-representation of pathways among subsets of the top N genes (ranked by association *P*-value) compared with all expressed genes. To ensure the robustness of this result, we repeated the analysis with varying N (N = {30, 60, 100, 150, 300, 500}). FDR was used for multiple hypothesis correction,[19] accounting for the number of assessed pathways.

Association of MDD with genetic variation and relationship with gene expression
Association between MDD and each genotyped SNP was evaluated using a standard logistic regression test, adjusted for five genotypic PCs (Supplementary Figure S4). We also looked up genotypic association *P*-values in the independent Psychiatric Genomics Consortium (PGC) MDD

Table 1. Covariate values for cases and controls

Covariate name	Cases	Controls	P-value
	Mean (s.d.)		(Case vs control)
Age at interview	44.77 (10.69)	44.64 (11.0)	NS
BMI	30.25 (7.74)	27.87 (6.45)	2.63E-07
Number of cigarettes per day	2.49 (6.76)	1.12 (4.55)	8.78E-06
Number of blood pressure meds	0.31 (0.66)	0.22 (0.53)	0.06
Current alcohol use	0.61 (0.78)	0.69 (0.76)	0.04
	N (%)		
N (cases or controls)	463	459	
Female gender	360 (0.778)	288 (0.627)	4.08E-07
Ate before blood draw	323 (0.698)	339 (0.739)	NS
Exercised before blood draw	74 (0.160)	92 (0.200)	0.064
Smoked before blood draw	67 (0.145)	34 (0.074)	4.01E-04
ACE inhibitor	36 (0.078)	30 (0.065)	NS
Anticholinergic meds	23 (0.050)	12 (0.026)	0.044
Antihistaminic meds	60 (0.130)	39 (0.085)	0.018
Beta blocker	44 (0.095)	14 (0.031)	3.45E-05
Cholesterol lowering meds	64 (0.138)	50 (0.109)	NS
Decongestant	41 (0.089)	10 (0.022)	4.94E-06
Decongestant or stimulant	52 (0.112)	10 (0.022)	1.27E-08
Diuretic	29 (0.063)	25 (0.054)	NS
Non-steroidal antiinflammatory	46 (0.099)	19 (0.041)	4.05E-04
Oral hypoglycemic	30 (0.065)	16 (0.035)	0.026
Oral contraceptive	45 (0.097)	39 (0.085)	NS
Protein pump inhibitor	52 (0.112)	32 (0.070)	0.016
Thyroid medication	54 (0.117)	34 (0.074)	0.018
Opiate medication or use	46 (0.099)	9 (0.020)	1.18E-07
Cannabis use (past 2 weeks)	23 (0.050)	9 (0.020)	9.61E-03

Abbreviations: ACE, angiotensin converting enzyme; BMI, body mass index; NS, not significant. The variables shown in this table were used as background in the likelihood ratio tests of association of major depressive disorder with expression levels of genes or pathways (Supplementary Figure S6); 10 expression principal components (PCs) and five genotypic PCs were also included. P-values were computed by Fisher's exact tests (binary variables) or Spearman's rank correlation tests (non-binary variables). Note that some of the medication categories are partially overlapping. Supplementary Table S2 shows levels of association of each of these variables (and of genotypic PCs) with expression PCs.

GWAS (9240 cases and 9519 controls).[3] In a joint analysis of gene expression and genotypes, we identified genes whose genetic effect on MDD may be mediated through altering gene expression. Here, we used as a test statistic the least significant of two P-values for association of MDD with (i) expression of that gene and (ii) genotype of its strongest eQTL SNP (Supplementary Methods). By evaluating the maximum (the least significant) of these two P-values as a single test statistic, this analysis identifies relationships in which both expression and genotype support the association between the gene and MDD. Significance was evaluated by computing this statistic for 1 000 000 permutations of MDD status for each gene. To derive stable estimates of P-values, a Weibull extreme value distribution was fit to the permutation test statistics and used as the null distribution to estimate the probability of observing the statistic seen in the real data.

Clinical variables

Factor scores (PCs analysis) were computed separately from clinical variables and childhood trauma questionnaire responses, and association of these scores with IFN α/β signaling pathway (PC1) scores (Figure 1 legend) was analyzed by analysis of variance, corrected for age and sex (Supplementary Methods).

## RESULTS

Association of single genes with MDD status

Analysis of single genes did not identify a genome-wide significant association ($P < 3.6E-6$, correcting for 13 857 genes). An excess of small P-values was observed (Supplementary Figure S7). We quantified this excess by estimating the proportion of true-positive tests ($\pi_1 = 0.13$),[23] which indicated that expression levels of many genes are modestly associated with MDD. In

addition, we performed a power analysis to estimate the theoretical effect size with LRTs given the current cohort size.[26] We estimated that we have 80% power to detect genome-wide significant P-values ($< 3.6E-6$) for individual genes with an expected estimated odds ratio of ≥ 1.6 (log fold change of ~0.5). This analysis suggests that, should there be a true association from blood expression profiles, the odds ratio of is likely < 1.6. At a less stringent threshold (0.25 FDR), there were 29 associated genes (Table 2), with biological functions including innate immune processes, vesicle trafficking, cell cycle regulation and splicing. The top two genes were MINOS1 (organization of mitochondrial inner membrane, $P < 5E-6$) and COPG (part of the Coatomer complex involved in vesicle trafficking between ER and Golgi, $P < 8E-6$).

Pathway enrichment analysis

In analyses of the enrichment of pathways among sets of top genes, all tested subsets of 60–500 genes were significantly enriched (0.05 FDR) for one pathway: IFN α/β signaling (REACTOME) (Figure 1, Supplementary Figure S8, Tables 3 and 4, and Supplementary Table S7). GSEA also identified only this pathway at 0.05 FDR by permutation tests (uncorrected $P = 2.5E-5$, Bonferroni-correct $P < 0.05$). Of the 64 genes in this pathway annotated in MSigDB,[24] 49 genes were adequately expressed here, of which 20 had a nominal $P < 0.05$ (Table 4). Overexpression was observed for 34 of the 49 genes including 19 of the 20 nominally significant genes (Table 4 and Supplementary Table S7). No clinical variable was observed to predict IFN pathway (PC1) scores in cases (see Figure 1 legend and Supplementary Data 1 for the list of top 10 pathways).
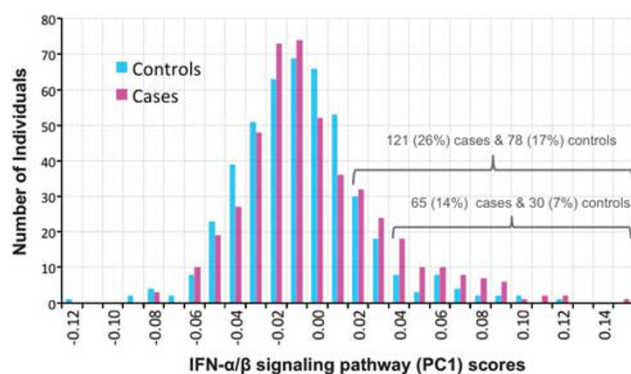
Figure 1. Interferon (IFN) α/β signaling pathway principal component (PC1) scores. Shown are the distributions of values of a score that summarizes expression levels of genes in the IFN α/β signaling pathway for cases (magenta) and controls (cyan). Each bar indicates the number of individuals with a score between the x axis value and the next higher value. Scores were computed as PC1 from PCs analysis of normalized read counts for the 20 genes (shown in Table 3) in the pathway with $P < 0.05$ for association with major depressive disorder (MDD) individually (among the 49 genes, of 64 in the pathway, that passed the inclusion criterion of ≥10 total reads in ≥100 individuals). Gene set enrichment analysis (GSEA) identified a significant association (0.05 FDR) between IFN α/β pathway gene expression and MDD (note that GSEA uses expression data for all genes, and not the summary PC1 score which is shown in this figure). There is an excess of cases with higher scores, as shown by the numbers over the brackets. Raw read counts were initially corrected for technical and biological covariates (Supplementary Table S1: specimen-specific sequencing variables, RNA quality, white cell-type proportion estimates, time of blood draw). Analysis of case-control difference included additional covariates (see Materials and methods) including medication and substance classes seen in ≥30 subjects (Table 2 and Supplementary Table S2). Case scores were not predicted by clinical variables or childhood trauma scores (Supplementary Results). Case-control differences and enrichment of top gene subsets for this pathway were not explained by excluding 91 individuals with rarer medical diagnoses or medications, estimating white cell-type proportions by a second method, controlling for intake of three antidepressant classes, or controlling for substance abuse/dependence or steroid medications.

Post-hoc re-assessment of potential confounders
Possible sources of spurious association with the IFN α/β signaling pathway were explored in *post hoc* analyses:

Unusual medication factors. We excluded 91 additional subjects with medications or illnesses with any potential impact on the immune system that were too rare for individual adjustment (Supplementary Table S4, for example, insulin, histamine-2 or leukotriene antagonists, antibiotics, immune suppressants; and autoimmune diagnoses). Enrichment of IFN α/β signaling pathway among subsets of 60–500 genes remained significant (0.05 FDR; Supplementary Table S3).

Cell-type proportions. Initial normalization of the data removed the effect of inferred cell-type proportions (Supplementary Methods), and as expected, we did not observe any correlation between these estimates and the residual expression levels of IFN pathway genes. (In raw data, IFN signaling is most strongly correlated with the proportion of activated dendritic cells, but that proportion is not predicted by MDD status.) As an additional check, we used an alternative computational method (ridge regression instead of non-negative least squares) to estimate cell-type proportions. The enrichment of IFN α/β signaling pathway among the top 100, 300 and 500 genes remained significant (0.05

Table 2. Top MDD-associated genes at 0.25 FDR

Gene Name	P-value	Direction	Function
MINOS1	5.00E-06	−	Mitochondrial component
COPG	8.00E-06	+	Vesicle trafficking (Golgi-to-ER)
SYF2	2.93E-05	−	Splicing factor
EHMT2	1.26E-04	+	Histone methylation
UBAP2L	1.26E-04	+	Transcription activation/acetylation
MAFK	1.26E-04	−	Transcription activation/repressor
MX1	1.26E-04	+	Immune function
C1ORF101	1.26E-04	+	(Not available)
SNRNP35	1.26E-04	−	Splicing factor
FAM120AOS	1.47E-04	−	(Not available)
NOP2	1.48E-04	+	Cell cycle regulation
SRSF5	1.64E-04	−	Splicing factor
TNFRSF10B	2.52E-04	+	Apoptosis, immune function
PIPOX	2.52E-04	+	(Not available)
OAS1	2.52E-04	+	Immune function
RABEPK	2.52E-04	+	Vesicle trafficking
SEMA3C	2.52E-04	+	Axon guidance
CINP	2.52E-04	−	Cell cycle regulation
H6PD	3.78E-03	−	ER, converting cortisone to cortisol
CCNI	3.78E-03	−	Cell cycle regulation
CROCC	3.78E-03	−	Cell cycle
IFIT3	3.78E-03	+	Immune function
MRFAP1	3.78E-03	−	Cell cycle regulation
STIP1	5.04E-04	+	Axon guidance
MORC3	5.04E-04	+	RNA binding
SDK1	5.04E-04	−	Axon guidance
RGPD1	5.04E-04	+	RNA binding
NXT1	5.04E-04	+	RNA binding and transport
PACSIN1	5.04E-04	+	Immune function, vesicle trafficking

Abbreviation: MDD, major depressive disorder. The table shows the top 29 genes (0.25 FDR) associated with MDD status. Direction indicates whether the association was with increased (+) or decreased (−) expression.

FDR), after accounting for re-estimated cell-type proportions in the LRT (Supplementary Results).

Antidepressants. In cases, we computed LRT *P*-values for association of all gene expression levels with three separate antidepressant classes taken by ≥30 individuals: serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, bupropion (Supplementary Table S3). Of the 20 nominally significant IFN signaling genes (Table 4), none was among the top 300 genes associated with any of these classes, and no enrichment for this pathway ($P>0.5$) was observed in sets of top 30–500 associated genes (whereas 17 of the top IFN signaling genes were among the top 300 in the primary analysis of association with MDD).

Drugs of abuse and steroids. Enrichment analyses were repeated (Supplementary Table S3) with additional covariates to account for substance dependence (derived from covariates, Supplementary Methods) and for Fagerstrom nicotine dependence score; and for classes of steroid medications (primarily inhaled steroids for asthma or allergies). IFN α/β signaling genes were enriched (0.05 FDR) in the subsets of top 100–500 genes, after accounting for these covariates in the LRT.

Association of genetic variation with MDD status
One common SNP showed genome-wide significant genotypic association with MDD (*rs11232553*, chr11:80,941,531; $P < 3E-8$; 1.8 Mb downstream of MIR708 and 660 kb upstream of MIR4300), but no association for this SNP was observed in the much larger PGC MDD data set (Supplementary Results). The joint analysis that combines expression data with genotype (Materials and methods) identified one genome-wide significant gene, CINP—0.05 FDR; uncorrected $P = 2.5E-4$ for expression level alone, 3.2e-04 for

Table 3. Enrichment P-values for association of IFN α/β signaling pathway and MDD

Pathway	N = 30	N = 60	N = 100	N = 150	N = 200	N = 300	N = 500
Interferon α/β signaling pathway	0.25 FDR	1.0E-06	7.0E-07	3.0E-11	2.0E-11	3.0E-13	3.0E-16

Abbreviations: IFN, interferon; MDD, major depressive disorder. Using the hypergeometric test, over-representation of pathways among the top N-associated genes was assessed for significance. The table shows results for the significantly over-represented pathway at 0.05 FDR (see Supplementary Table S7 for complete list of genes and single-gene association P-values).

Table 4. IFN α/β signaling pathway genes with the strongest associations with MDD

Gene Name	P-value	Rank among all genes	Direction
MX1	1.26E-04	7	+
OAS1	2.52E-04	15	+
IFIT3	3.78E-04	22	+
PTPN6	1.26E-03	43	+
ADAR	1.26E-03	45	+
IRF7	1.64E-03	53	+
IFIT1	3.15E-03	100	+
USP18	3.78E-03	121	+
ISG15	3.78E-03	122	+
OAS2	4.16E-03	127	+
IRF8	6.93E-03	185	+
IFIT2	7.30E-03	201	+
OAS3	7.56E-03	208	+
MX2	8.44E-03	226	+
IRF9	1.07E-02	274	+
IFI6	1.13E-02	285	+
OASL	1.44E-02	352	+
XAF1	1.49E-02	362	+
IFI35	1.65E-02	404	+
IFNAR2	4.99E-02	1030	−

Abbreviations: IFN, interferon; MDD, major depressive disorder. The table shows the top twenty genes (those associated with nominal P-value < 0.05) in the interferon α/β signaling pathway, along with P-values, direction of effect, and association rank among all expressed genes.

genotypic association of its top eQTL (rs2896439), 1e-6 for the combined test. A modest association of rs2896439 in the PGC cohort (P = 0.035) suggests caution in interpreting our result. Three other genes achieved FDR < 0.1: SCAI (cell migration and regulation of cell cycle), SDK1 (axon guidance and implicated in HIV-associated nephropathy) and RABEPK (endosome to Golgi trafficking). We did not detect evidence of a relationship between genetic variation affecting IFN pathway genes and MDD in either analysis (joint analysis or genotype-only) of this cohort, or in a targeted analysis of the PGC data set (Supplementary Results).

DISCUSSION

Type I IFN signaling and MDD

Significant association was observed between MDD and expression of IFN α/β signaling pathway genes in a large population-based sample, using whole-blood RNA reflecting the physiological state at the time of blood draw. Primary analyses removed the effects of a large set of technical, biological, medical and hidden covariates, and post hoc analyses did not identify measured factors that explained the finding.

IFN-α and IFN-β are type I IFNs (IFN-I), the main cytokines of the innate immune system that respond primarily to viral infection and to malignant cells. They activate genes that interfere with viral replication, activate other immune responses to infection and inhibit cell growth.[27] A 'weak' (chronic) signaling mode, that is

observed in the absence of known pathogens, may increase the efficiency of response to stimuli, but may also have a role in autoimmune and neuroinflammatory disorders.[28] The binding of IFN-I to IFN receptors (IFNAR) can activate two main transcriptional complexes: IFN-α-activated factor, which mainly mediates activation of other cytokines, and IFN-stimulated gene factor 3 (ISGF3), which mainly mediates antiviral activity. Here, we observed the upregulation of ISGF3 induced genes in MDD cases (Table 4), including IRF9, a component of the ISGF3 complex.[29]

Increased IFN-I signaling in MDD is consistent with previous data implicating dysregulation of cytokines in depression. The most direct observation is that patients receiving IFN-I therapies (IFN-α for hepatitis C or IFN-β for multiple sclerosis) often develop clinically significant depression.[30–33] Previous studies have associated changes in secreted cytokines and inflammatory markers, including IFN-I,[34,35] with a reduction in tryptophan, the precursor of serotonin, in both cerebral spinal fluid[35] and plasma.[36]

More broadly, increased IFN-I signaling is relevant to a set of inter-related findings regarding the role of immune system in the pathophysiology of MDD (reviewed by Zunszain et al.[37]) (i) glucocorticoid dysregulation: hypersecretion of cortisol (stimulated by corticotropin releasing factor) is observed in MDD, possibly caused or mediated by glucocorticoid receptor resistance.[37] (ii) Immune/inflammatory dysregulation: levels of circulating inflammatory cytokines are increased, particularly tumor necrosis factor alpha (TNF-α) and interleukin 6 (IL-6).[38–40] Also, plasma levels of a non-specific inflammation marker, C-reactive protein, predicted depression in a Danish sample of over 73,000 individuals.[41] (iii) Hippocampal volume is reduced with longer durations of MDD,[42] due either to cell loss (implicating inflammation-related apoptosis[37,43]) or to reduced cellular volume.[37]

Existing studies suggest multiple hypotheses concerning the causal relationships among these findings. An initiating factor could be stress-induced glucocorticoid secretion with resulting neuroinflammation, cytokine production and apoptosis.[44] Alternatively, dysregulated cytokine production could directly cause depression (perhaps mediated by tryptophan depletion) and also increase glucocorticoid receptor resistance.[45] Indeed, it has been proposed that sequence variants that predispose to depression in modern society may have been useful ancient adaptations to pathogens.[46] It is reasonable to assume that there are multiple causal inter-relationships, reflecting a high degree of phenotypic and etiologic heterogeneity in MDD.

The finding of increased expression levels of IFN-I signaling pathway in this study is consistent with all of the above hypotheses. It could represent a key causal or mediating factor, and/or the end result of a complex process. However, a third possibility is that increased expression of IFN-I signaling is caused by a confounding factor that is downstream of or simply correlated with MDD. We did not find genetic variants impacting IFN pathway genes (either protein-coding variants or known eQTLs) to be significantly associated with MDD in this cohort. We note that there are multiple IFN-I gene subtypes in a cluster on chromosome 9, each of which had too few uniquely-mapped reads to meet the criterion for inclusion in the analyses for this study. They have highly homologous sequences, so that many

sequencing reads in this region were excluded here because we counted only uniquely mapped reads.

We also should not overlook the possibility that, as the primary function of the *IFN-I* signaling system is response to viral infection, a subset of MDD patients could have unrecognized chronic and/or active infections, whose effects could be influenced by many genetic sequence variants and by other environmental factors. This hypothesis cannot be directly addressed by the present data but suggests an interesting future research direction.

Comparison with previous studies

MDD transcriptome studies of postmortem brain tissue have suggested alterations in glutamatergic and GABA-ergic pathways[47] and in synaptic genes relevant to the anatomical finding of reduced dendritic spines.[48] Interpretation is complicated by confounding factors (postmortem changes and drug exposures) and small samples (typically fewer than 30 cases).

Previous blood cell gene expression studies—typically targeting 1–15 genes in samples of 25–100 MDD cases—have reported expression increases for cytokine- and glucocorticoid-related genes, and reductions for glucocorticoid receptor and neuroplasticity-related genes such as *BDNF*.[49] One study reported increased blood expression levels of *IFN-α1*, *IFN-α2* and *IFN-β1* in 22 MDD cases and 11 controls.[32] There are two reports from a larger study that included microarray whole-transcriptome and GWAS data for 215 MDD cases and their relatives (San Antonio Family Study): *RNF123* was the most significant gene in a bivariate linkage analysis of MDD and gene expression values ($P > 0.5$ here);[50] and *DISC1* SNPs showed genotypic association to MDD[51] ($P = 0.025$ for *DISC1* here, but with increased expression, whereas previous studies predict reduced expression of *DISC1*). Whole-transcriptome association results have not been reported to date for that study. Interestingly, a recent study of subjective social isolation reported dysregulation of pro- and antiinflammatory genes, including downregulation of IFN I signaling pathway in lonely individuals;[52,53] we note that the impact of other unmeasured factors that modulate the IFN response (for example, age differences in the cohorts) may underlie the apparent discrepancy in the direction of effect of IFN I signaling.

Meta-analysis of studies of serum cytokine levels in MDD suggests significant increases in *TNF-α* and *IL-6*;[38] here, overexpression of *TNFRSF10B* was the thirteenth strongest individual gene finding (Table 2), whereas overexpression of *IL-6* was not observed, although the strongest *IL-6* findings have been in cerebral spinal fluid.[35] There have been studies of *IFN-λ* (without significant findings overall[38]), but not type I IFNs. We were unable to measure serum biomarkers in the present study because the study was designed to collect a large national sample, and this design did not permit obtaining and then immediately centrifuging and preserving serum.

CINP and IFN

Joint analysis of association of MDD to expression levels and to genotypes detected an association with *CINP*. The antiviral component of IFN signaling results in cell cycle arrest and inhibition of cell growth by inhibiting cyclin-dependent kinases,[9] and specifically *CDK2*.[54] *CINP* interacts with *CDK2* as a component of the DNA damage integrity checkpoint, and its suppression has been linked to cell cycle arrest.[8] As noted above, however, support for this association is very modest in the larger PGC MDD cohort.

Strengths and limitations

Strengths of this study design include deep RNA-seq data; a large population-based sample with a representative female:male ratio (2:1), including medicated and unmedicated cases; a relevant tissue type for immunological mechanisms; and correction for multiple measured and unmeasured covariates. Limitations include use of non-brain tissue; the use of mixed white blood cell types (as opposed to cell-sorted data) and not having cell count measurements; and case-control differences in many physiological covariates (Table 1 and Supplementary Table S2). The detection of a single associated pathway, despite an excess of modestly significant P-values, also suggests that (as in GWAS) larger samples are needed. Our reliance on log-linear models to account for covariates could have left residual correlation due to non-linear effects. Also, we only adjusted for covariates relevant to at least 30 subjects in the primary analysis. Under-reporting of medical information by cases could have produced spurious results given our reliance on self-report (although none of those variables were associated with IFN-I signaling). IFN-I signaling could have been influenced by common viral infections unrelated to MDD, although these are seasonal and case status was not related to order of blood collection. We used two statistical estimates of white cell-type proportions based on expression signatures from microarray data, but these estimates might not have been sufficiently accurate, and they were based on microarray measurements, whereas our data is based on RNA-seq.

Finally, results were partially dependent on our approach to removing hidden confounding factors by accounting for expression PCs, which has been shown to produce more accurate estimations of eQTLs[21,55] and of functional gene interactions.[22] To avoid over-fitting, we removed the top 10 PCs based only on explained variance (Supplementary Figure S6), ignoring MDD status.

In a *post hoc* assessment, we do observe weaker (non-significant) associations between the IFN pathway and MDD status without removing the effect of top expression PCs. Although this indicates some sensitivity of our analysis to modeling choices, we have observed that the top expression PCs are correlated with confounding factors including age (Supplementary Figure S9), and thus it is likely that the chosen procedure yields more meaningful results.

CONCLUSION

We identified a significant association between MDD and increased expression of genes involved in IFN α/β signaling. This supports hypotheses of dysregulated cytokine activity in MDD, but gene expression data cannot resolve critical cause–effect relationships. Increased *IFN-I* signaling could be a direct cause of depression, resulting from some combination of genetic sequence variants, psychological stress, or normal or abnormal responses to unknown viral infection or other physiological stressors. Increased signaling could also be downstream of depression, or simply based on correlation of *IFN-I* signaling with unknown confounding variables that are correlated with depression. There are also plausible hypotheses (with several possible directionalities of effects) involving an interaction between immune dysregulation and increased resistance to the release of glucocorticoids. There is a need for studies that simultaneously measure clinical variables, genotypes, gene expression, viral sequences or antibodies, and proteins related to immune function and glucocorticoid dynamics.

Data access

Data from this study are available to qualified scientists from the NIMH Center for Collaborative Genomic Studies on Mental Disorders, including raw and aligned RNA-seq data, raw and normalized read counts per gene, and covariate data. Scientists who are interested in requesting data should consult https://www.nimhgenetics.org/access_data_biomaterial.php, which provides specific instructions about how to request data, including the name and e-mail address of the program official who is responsible for handling requests. When inquiring about data access, please reference this study as the Depression Genes and Networks study (DFL, PI).

REFERENCES

1 Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 2000; **157**: 1552–1562.

2 Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. *Diabetologia* 1999; **42**: 139–145.

3 Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, Breen G *et al*. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* 2013; **18**: 497–511.

4 Rush AJ. The varied clinical presentations of major depressive disorder. *J Clin Psychiatry* 2007; **68**(Suppl 8): 4–10.

5 Sullivan PF, de Geus EJ, Willemsen G, James MR, Smit JH, Zandbelt T *et al*. Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry* 2009; **14**: 359–375.

6 Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, Breen G *et al*. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* 2013; **18**: 497–511.

7 Klengel T, Binder EB. Gene-environment interactions in major depressive disorder. *Can J Psychiatry* 2013; **58**: 76–83.

8 Grishina I, Lattes B. A novel Cdk2 interactor is phosphorylated by Cdc7 and associates with components of the replication complexes. *Cell Cycle* 2005; **4**: 1120–1126.

9 Sangfelt O, Erickson S, Castro J, Heiden T, Gustafsson A, Einhorn S *et al*. Molecular mechanisms underlying interferon-alpha-induced G0/G1 arrest: CKI-mediated regulation of G1 Cdk-complexes and activation of pocket proteins. *Oncogene* 1999; **18**: 2798–2810.

10 Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MW, McCormick C *et al*. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Research* 2013; doi:10.1101/gr.155192.113.

11 Kessler RC, Andrews G, Mroczek D, Ustun B, Wittchen H. The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF). *Int J Methods Psychiatric Res* 1998; **7**: 171–185.

12 Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Int Medi* 2001; **16**: 606–613.

13 Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Int Medi* 2006; **166**: 1092–1097.

14 Debey-Pascher S, Hofmann A, Kreusch F, Schuler G, Schuler-Thurner B, Schultze JL *et al*. RNA-stabilized whole blood samples but not peripheral blood mononuclear cells can be stored for prolonged time periods prior to transcriptome analysis. *J Mol Diagn* 2011; **13**: 452–460.

15 Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; **25**: 1105–1111.

16 Anders S HTSeq: Analysing high-throughput sequencing data with Python http://www-huberembhde/users/anders/HTSeq/doc/overviewhtml. 2010.

17 Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z, Clark HF. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PloS ONE* 2009; **4**: e6098.

18 Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F *et al*. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun* 2005; **6**: 319–331.

19 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc* 1995; **57**: 289–300.

20 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet* 2006; **38**: 904–909.

21 Stegle O, Parts L, Durbin R, Winn J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol* 2010; **6**: e1000770.

22 Furlotte NA, Kang HM, Ye C, Eskin E. Mixed-model coexpression: calculating gene coexpression while accounting for expression heterogeneity. *Bioinformatics* 2011; **27**: i288–i294.

23 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003; **100**: 9440–9445.

24 Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011; **27**: 1739–1740.

25 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Nati Acad Sci USA* 2005; **102**: 15545–15550.

26 Shieh G. On power and sample size calculations for likelihood ratio tests in generalized linear models. *Biometrics* 2000; **56**: 1192–1196.

27 Platanias LC. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat Rev Immunol* 2005; **5**: 375–386.

28 Taniguchi T, Takaoka A. A weak signal for strong responses: interferon-alpha/beta revisited. *Nat Rev Mol Cell Biol* 2001; **2**: 378–386.

29 Wang BX, Fish EN. The yin and yang of viruses and interferons. *Trends Immunol* 2012; **33**: 190–197.

30 Udina M, Castellvi P, Moreno-Espana J, Navines R, Valdes M, Forns X *et al*. Interferon-induced depression in chronic hepatitis C: a systematic review and meta-analysis. *J Clin Psychiatry* 2012; **73**: 1128–1138.

31 Schaefer M, Engelbrecht MA, Gut O, Fiebich BL, Bauer J, Schmidt F *et al*. Interferon alpha (IFNalpha) and psychiatric syndromes: a review. *Prog Neuro-psychopharmacol Biol Psychiatry* 2002; **26**: 731–746.

32 Schlaak JF, Trippler M, Hoyo-Becerra C, Erim Y, Kis B, Wang B *et al*. Selective hyper-responsiveness of the interferon system in major depressive disorders and depression induced by interferon therapy. *PloS ONE* 2012; **7**: e38668.

33 Felger JC, Cole SW, Pace TW, Hu F, Woolwine BJ, Doho GH *et al*. Molecular signatures of peripheral blood mononuclear cells during chronic interferon-alpha treatment: relationship with depression and fatigue. *Psychological Med* 2012; **42**: 1591–1603.

34 Raison CL, Borisov AS, Woolwine BJ, Massung B, Vogt G, Miller AH. Interferon-alpha effects on diurnal hypothalamic-pituitary-adrenal axis activity: relationship with proinflammatory cytokines and behavior. *Mol Psychiatry* 2010; **15**: 535–547.

35 Raison CL, Borisov AS, Majer M, Drake DF, Pagnoni G, Woolwine BJ *et al*. Activation of central nervous system inflammatory pathways by interferon-alpha: relationship to monoamines and depression. *Biol Psychiatry* 2009; **65**: 296–303.

36 Hughes MM, Carballedo A, McLoughlin DM, Amico F, Harkin A, Frodl T *et al*. Tryptophan depletion in depressed patients occurs independent of kynurenine pathway activation. *Brain Behav Immun* 2012; **26**: 979–987.

37 Zunszain PA, Anacker C, Cattaneo A, Carvalho LA, Pariante CM. Glucocorticoids, cytokines and brain abnormalities in depression. *Prog Neuro-psychopharmacol Biol Psychiatry* 2011; **35**: 722–729.

38 Dowlati Y, Herrmann N, Swardfager W, Liu H, Sham L, Reim EK *et al*. A meta-analysis of cytokines in major depression. *Biol Psychiatry* 2010; **67**: 446–457.

39 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; **9**: 46–56.

40 Miller AH, Maletic V, Raison CL. Inflammation and its discontents: the role of cytokines in the pathophysiology of major depression. *Biol Psychiatry* 2009; **65**: 732–741.

41 Wium-Andersen MK, Orsted DD, Nielsen SF, Nordestgaard BG. Elevated C-reactive protein levels, psychological distress, and depression in 73, 131 individuals. *JAMA Psychiatry* 2013; **70**: 176–184.

42 McKinnon MC, Yucel K, Nazarov A, MacQueen GM. A meta-analysis examining clinical predictors of hippocampal volume in patients with major depressive disorder. *J Psychiatry Neurosci* 2009; **34**: 41–54.

43 Nadeau S, Rivest S. Glucocorticoids play a fundamental role in protecting the brain during innate immune response. *J Neurosci* 2003; **23**: 5536–5544.

44 Heim C, Newport DJ, Mletzko T, Miller AH, Nemeroff CB. The link between childhood trauma and depression: insights from HPA axis studies in humans. *Psychoneuroendocrinology* 2008; **33**: 693–710.

45 Leonard B, Maes M. Mechanistic explanations how cell-mediated immune activation, inflammation and oxidative and nitrosative stress pathways and their sequels and concomitants play a role in the pathophysiology of unipolar depression. *Neurosci Biobehav Rev* 2012; **36**: 764–785.

46 Raison CL, Miller AH. The evolutionary significance of depression in Pathogen Host Defense (PATHOS-D). *Mol Psychiatry* 2013; **18**: 15–37.

47 Sequeira A, Mamdani F, Ernst C, Vawter MP, Bunney WE, Lebel V *et al*. Global brain gene expression analysis links glutamatergic and GABAergic alterations to suicide and major depression. *PloS ONE* 2009; **4**: e6585.

48 Kang HJ, Voleti B, Hajszan T, Rajkowska G, Stockmeier CA, Licznerski P *et al*. Decreased expression of synapse-related genes and loss of synapses in major depressive disorder. *Nat Med* 2012; **18**: 1413–1417.

49 Hepgul N, Cattaneo A, Zunszain PA, Pariante CM. Depression pathogenesis and treatment: what can we learn from blood mRNA expression? *BMC Med* 2013; **11**: 28.

50 Glahn DC, Curran JE, Winkler AM, Carless MA, Kent JW Jr., Charlesworth JC *et al.* High dimensional endophenotype ranking in the search for major depression risk genes. *Biol Psychiatry* 2012; **71**: 6–14.

51 Carless MA, Glahn DC, Johnson MP, Curran JE, Bozaoglu K, Dyer TD *et al.* Impact of DISC1 variation on neuroanatomical and neurocognitive phenotypes. *Mol Psychiatry* 2011; **16**, 1096–1104.

52 Cole SW, Hawkley LC, Arevalo JM, Cacioppo JT. Transcript origin analysis identifies antigen-presenting cells as primary targets of socially regulated gene expression in leukocytes. *Proc Natl Acad Sci USA* 2011; **108**: 3080–3085.

53 Cole SW, Hawkley LC, Arevalo JM, Sung CY, Rose RM, Cacioppo JT. Social regulation of gene expression in human leukocytes. *Genome Biol* 2007; **8**: R189.

54 Xie RL, Gupta S, Miele A, Shiffman D, Stein JL, Stein GS *et al.* The tumor suppressor interferon regulatory factor 1 interferes with SP1 activation to repress the human CDK2 promoter. *J Biol Chem* 2003; **278**: 26589–26596.

55 Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet* 2011; **7**: e1002003.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)