MAJOR ARTICLE

Peripheral Blood Gene Expression in Postinfective Fatigue Syndrome Following From Three Different Triggering Infections

Sally Galbraith,[1,a] Barbara Cameron,[2,a] Hui Li,[2] Diana Lau,[3] Ute Vollmer-Conna,[4] and Andrew R. Lloyd[2]

[1]School of Mathematics and Statistics, Faculty of Science, [2]Inflammation and Infection Research Centre, School of Medical Sciences, [3]Children's Cancer Research Institute, and [4]School of Psychiatry, Faculty of Medicine, University of New South Wales, Sydney, Australia

Background. Several infections trigger postinfective fatigue syndromes, which share key illness characteristics with each other and with chronic fatigue syndrome (CFS). Previous cross-sectional case-control studies of CFS have suggested that unique gene expression signatures are evident in peripheral blood samples.

Methods. Peripheral blood transcriptomes in samples collected longitudinally, in 18 subjects with a fatigue syndrome lasting ≥6 months after acute infection due to Epstein-Barr virus, Ross River virus, or *Coxiella burnetii* (Q fever), and 18 matched control subjects who had recovered promptly, were studied by microarray (n = 127) and confirmatory quantitative polymerase chain reaction (PCR). Gene expression patterns associated with CFS were sought by univariate statistics and regression modeling.

Results. There were 23 genes with modest differential expression (0.6–2.3-fold change) in within-subject comparisons of early, symptomatic time points with late, recovered time points. There were modest differences found in 63 genes, either in cross-sectional comparison of cases and controls at 6 months after infection onset or in the regression model. There were 223 genes significantly correlated with individual symptom domains. Quantitative PCR confirmed 33 (73%) of 45 genes—none were consistent across cohorts.

Conclusions. Although the illness characteristics of patients with postinfective fatigue syndromes have more similarities than differences, no reliable peripheral blood gene expression correlate is evident.

The postinfective fatigue syndrome is the only validated model for studies of the onset and evolution of chronic fatigue syndrome (CFS) [1–4]. The pathophysiology of CFS remains unknown [5, 6], although the link with acute infection has led to exhaustive, and essentially negative, investigations seeking persistent pathogens or immune dysfunction [2, 7, 8].

This impasse has driven the search for novel biomarkers, including alterations in leukocyte gene expression in case-control series of patients with CFS and healthy subjects [9–12]. A series of putatively correlated genes have been identified, but none were shared between studies, raising concerns of false discovery. Zhang and colleagues recently reported replication of a gene expression "signature" in a second case-control study, although perplexingly, the first [12] and second case groups [13] were merged for the analysis. Importantly, a recent study of the peripheral blood transcriptome in 44 monozygotic twin pairs discordant for chronic fatigue revealed no evidence of differential gene expression [14]. This study provided the ideal control for the potential confounding effects of mismatched genetic backgrounds in the earlier case-control studies.

Another potential explanation for nonreproducible results may relate to the recognized heterogeneity within the label of CFS [15–17]. Four approaches have been adopted to overcome this: first, identification of subgroups within large case series with homogeneous illness characteristics and unique signatures. A series of reports [18–20] from one data set [21] adopted this approach; however, these analyses were post hoc after the results of the primary case-control comparison were negative.

Received 26 February 2011; accepted 5 July 2011; electronically published 29 September 2011.

[a]These authors contributed equally to this study.

Correspondence: Barbara Cameron, Phd, Inflammation and Infection Research Centre, School of Medical Sciences, University of New South Wales, Sydney 2052, Australia (b.Cameron@unsw.edu.au).

The Journal of Infectious Diseases 2011;204:1632–40

0022-1899 (print)/1537-6613 (online)/2011/20410-0021$14.00
DOI: 10.1093/infdis/jir612

A second approach has been empirical delineation of subgroups within case series by application of cluster analysis (or similar) to the gene expression data set, followed by consideration of potentially coherent clinical features of the subgroups [22–25]. These observations await the critical test of independent replication. The third approach has been to seek gene expression correlates of ''endophenotypes,'' or coherent symptom domains within the label of CFS derived by factor, or principal components, analysis [26]. This approach is analogous to investigations of the common, complex disorders of mental health, including major depression, in which the behavioral symptom complex is divided into recognizable and stable characteristics with a likely biological association [27]. Again, the findings in CFS await independent replication.

The final approach has been to investigate potentially homogeneous subject groups, such as postinfective fatigue syndromes. One cross-sectional study of postinfective fatigue syndrome case subjects and healthy subjects has been reported, with differential expression of only a single gene (for the chemokine receptor CXCR4) being shared with those previously reported in CFS [28]. We reported leukocyte gene expression in longitudinally collected samples from case subjects with a postinfective fatigue syndrome after documented acute infectious mononucleosis and matched control subjects who recovered uneventfully after the same infection [29]—an approach which offers the additional methodological strength of comparison of samples obtained during the illness with samples from the same individual upon recovery. No differentially expressed genes shared with those in any previous studies were found.

Based on an assumption that a central premise of the syndromal diagnosis of CFS is that subjects share key clinical features, and hence gene expression correlates, we report here the findings of an analysis of the peripheral blood transcriptome in 3 longitudinal, nested case-control series of subjects with postinfective fatigue syndrome and matched individuals who recovered uneventfully following from the infective triggers. We sought differential expression between early illness and late recovery, postinfective fatigue syndrome case subjects and matched recovered control subjects, and genes that were correlated with endophenotypes derived by principal components analysis.

SUBJECTS AND METHODS

Subjects

Participants were white individuals enrolled in the Dubbo Infection Outcomes Study (DIOS) cohort, following confirmation of acute Epstein-Barr virus (EBV), Ross River virus (RRV), or *Coxiella burnetii* (Q fever) infection [30]. At each visit, detailed self-report and interview assessments of physical and psychological health were recorded. The severity and duration of symptoms were monitored using the SPHERE questionnaire [31], in which a validated Somatic and Physical Health Report questionnaire Somatic items (SOMA) subscale records the key outcome—that is, a score of $\geq$3 represents a clinically-significant fatigue state [32, 33].

In those subjects with persistent symptoms for $\geq$3 months, structured medical and psychiatric assessments were undertaken to exclude unrelated causes of illness. Eighteen subjects with postinfective fatigue syndromes (ie, those who had unexplained illness persisting for $>$6 months after onset of symptoms, meeting diagnostic criteria for CFS) and 18 control subjects who recovered promptly, matched by age, sex, and infection type, were selected [2] (Table 1). To allow investigation of the gene expression correlates of key components of the symptom complex, principal components analysis–derived indices [2] for 4 symptom domains (fatigue, pain, mood disturbance, and neurocognitive disturbance) and overall illness severity were calculated for each subject and time point from the self-report data sets.

The study protocol was approved by the relevant institutional review boards. All subjects provided written informed consent.

Specimens and Laboratory Methods

Blood samples were collected in the morning and transported to the laboratory within 6 hours, and then peripheral blood mononuclear cells (PBMCs) were separated (Lymphoprep; Axis-Shield, Oslo, Norway) and cryopreserved in vapor-phase liquid nitrogen. Thawed PBMCs were lysed in Tri-reagent (Sigma-Aldrich), RNA was extracted and quantified, and the quality was evaluated by denaturing gel electrophoresis and Bioanalyzer.

A total of 127 samples were analyzed, including 3 or 4 time points per subject (Table 1). RNA of sufficient yield and high quality (28S:18S ratio, 1.8–2.0) was available for all samples. For each subject, probe synthesis and hybridizations for all samples were performed in a single run. Arrays from a case and a control subject were run together whenever possible to control for run-to-run effects. For each sample, 500 ng of RNA was amplified to biotin-labeled complementary RNA (Illumina TotalPrep RNA amplification kit; Ambion, Austin, TX) following manufacturer's instructions.

Microarrays

Sentrix HumanRef-8 v2 Expression BeadChip microarrays (Illumina, San Diego, CA) were utilized. Each BeadChip carries 8 arrays with 22 184 probes identifying 18 203 unique genes. The microarray procedure was performed as instructed by the manufacturer. All reagents and buffers were provided with the BeadChip kits (Illumina). The arrays were stained with streptavidin-Cy3 (1 mg/mL; Sigma-Aldrich), and the bar-coded chips were scanned (Illumina) using the recommended settings. Images were analyzed with Beadstudio Gene Expression Module software (Illumina) to provide raw expression data.

Table 1. Subject Groups, Symptom Scores, and Sampling Time Points for the Microarray Analysis

				Symptom scores[a]				
				T1	T2	T3		T4
Subject[b]	Sex	Age	Infection	0 < 6 wks	6 < 12 wks	3 < 9 mths	>9 mths	>12 mths
PIFS1[†]	F	18	EBV	4	7	8	…	6
PIFS2	M	48	EBV	6	5	4	…	4
PIFS3	F	63	RRV	…	5	10	…	7
PIFS4	F	42	RRV	8	…	6	6	4
PIFS5	M	62	QF	8	11	12	…	9
PIFS6	M	36	QF	…	6	4[c], 10	…	9
PIFS7	M	68	QF	8	…	10	…	5
PIFS8	F	16	EBV	9	11	10	…	7
PIFS9	M	24	RRV	8	5	3	…	11
PIFS10	F	16	EBV	7	8	8	…	7
PIFS11	M	49	QF	9	8	6	…	6
PIFS12	M	42	RRV	…	9	4[c]	10	0
PIFS13	M	25	QF	10	12	12[c]	6	…
PIFS14	F	20	EBV	…	7	6[c]	5	1
PIFS15	M	60	RRV	11	8[c], 10	…	…	2
PIFS16	F	63	QF	5	5	4	…	0
PIFS17	M	35	RRV	10	8	5	…	…
PIFS18	M	29	EBV	…	3	3[c]	1	1
C1	M	16	EBV	8	5	2	…	2
C2	F	20	EBV	4	7	…	…	…
C3	F	32	RRV	8	6	…	…	0
C4	M	42	QF	8	3	…	…	…
C5	F	47	QF	6	3	…	…	1
C6	F	23	EBV	4	1	…	…	0
C7	M	38	QF	11[c], 4	…	…	…	2
C8	F	22	EBV	6	0	…	…	…
C9	F	20	EBV	7	2	…	…	0
C10	M	44	EBV	6	2	…	…	…
C11	F	40	RRV	6	0	…	1	0
C12	M	53	RRV	11	0	…	…	…
C13	M	37	RRV	4	0	…	…	0
C14	M	69	RRV	3	1	…	…	…
C15	M	47	QF	4	1	…	…	1
C16	M	70	QF	4	1	…	…	1
C17	M	51	QF	4	0	…	…	…
C18	M	30	RRV	0	2	…	…	1

Abbreviations: EBV, Epstein-Barr virus; F, female; M, male; QF, Q fever (*Coxiella burnetii*); RRV, Ross River virus; T1, time point 1; T2, time point 2; T3, time point 3; T4, time point 4.

[a] Symptom scores on the SOMA subscale of the SPHERE questionnaire (possible range, 0–12); a score of ≥3 indicates a clinically significant fatigue state.

[b] PIFS refers to case subjects with postinfective fatigue syndrome; C refers to control subjects who recovered within 6 weeks of onset.

[c] Where there are 2 arrays for the same time point, the array corresponding to the earlier measured SOMA score was used in the analysis.

Gene Expression Validation

The HighCapacity RNA to complementary DNA (cDNA) kit (Applied Biosystems, Carlsbad, CA) was used to convert messenger RNA to cDNA, before freezing at −20°C until use. Confirmatory gene expression was measured using Taqman assays formatted onto microfluidic cards in a 384-well format (Applied Biosystems 7900HT Fast Real-Time polymerase chain reaction (PCR) system). The assay primers were prevalidated by the manufacturer and selected to exclude genomic DNA amplification. The cDNA was mixed with TaqMan Universal PCR Master mix and loaded into the fill ports (500 ng of cDNA per port). Three housekeeping genes with stable expression in peripheral blood leukocytes were included (glyceraldehyde 3-phosphate dehydrogenase [*GAPDH*], ubiquitin C [*UBC*], and tyrosine monooxygenase [*YWHAZ*]) [34]. Cycle threshold (Ct) values were determined using 7900HT software before values were collated for analysis (DataAssist software, version 2.0; Applied Biosystems).

Statistical Analyses

Normalization

Raw array expression levels were $\log_2$ transformed and quantile normalized. For the paired analysis, arrays for each subject were normalized together. For the 2-sample analysis, the 28 arrays included were normalized together; and for the longitudinal and correlation analyses, all 127 arrays were normalized together.

When 2 arrays were available for any subject time point, the data for the array associated with the earlier date of collection was utilized. From each analysis, genes were ranked according to *P* value separately for EBV, RRV, and Q fever, and those genes in the top 2000 for all infections that were consistently up- or down-regulated were identified.

Paired Analysis

Within-subject comparisons were used to find genes differentially expressed between early illness (termed here "sick") and late recovery (termed here "well"). Twelve subjects (3 with EBV infection, 4 with RRV infection, and 5 with Q fever) (Table 1) with SOMA scores of ≥3 at time point 1 (T1) and SOMA scores of <3 at time point 4 (T4) were identified. Paired *t* tests were used to test for differences in normalized expression for each gene. Four sets of tests were performed: 1 for the whole group of 12 subjects, and 1 for each of the 3 infection types. Moderated *t* statistics, with unadjusted and Benjamini-Hochberg adjusted *P* values, were obtained in R software (http://www.R-project.org) using the *Bioconductor* package limma [35].

Two-sample Analysis

Between-subject comparisons were used to identify genes differentially expressed between case subjects with a postinfective fatigue syndrome and recovered subjects at the 6-month time point (T3). This included 17 case subjects (6 with EBV infection, 5 with RRV infection, and 6 with Q fever) (Table 1) and 11 control subjects (5 with EBV infection, 4 with RRV infection, and 2 with Q fever) (Table 1). Two-sample *t* tests for each gene were used to test for a difference in normalized expression levels between case subjects and recovered control subjects. Tests were performed using all arrays, as well as separately for the 3 infection subgroups. Moderated *t* statistics, with unadjusted and Benjamini-Hochberg adjusted *P* values, were obtained.

Longitudinal Analysis

To investigate differences in gene expression between subjects with postinfective fatigue (sick) and recovered status (well) over the whole time course, the following logistic regression model was used:

$$\begin{aligned}\log\left(\frac{p}{1-p}\right) = {} & \beta_0 + \beta_1 \text{age} + \beta_2 \text{sex} + \beta_3 \text{QF} + \beta_4 \text{RRV} + \beta_5 \text{T2} \\ & + \beta_6 \text{T3} + \beta_7 \text{T4} + \beta_8 \text{gene} \\ & + \beta_9 \text{QF} \times \text{gene} + \beta_{10} \text{RRV} \times \text{gene}\end{aligned}$$

where *P* represents the probability of experiencing fatigue (SOMA score, ≥3 or more), "sex" is an indicator for gender, "Q fever" and "RRV" are indicators for Q fever and RRV, respectively (EBV is the reference category), "gene" is gene expression, and T2, T3, and T4 are indicators for time points 2, 3, and 4, respectively (T1 is the reference category). The last 2 terms allowed for gene by disease interaction. The model was fitted for each gene using generalized estimating equations (GEEs) with working independence, with subject as cluster. All subjects and time points were included in this analysis. The analysis was performed in R using the package *geepack* [36] to obtain *P* values (unadjusted and Benjamini-Hochberg adjusted) for differentially expressed genes separately for the 3 infective cohorts. Because the parameters in the model have interpretations as log-odds ratios, tests for differential expression between sick and well states were formulated by testing whether the relevant parameters were equal to 0. For example, a test of the null hypothesis $\beta_8 = 0$ can be used to assess whether gene expression has an effect on the probability of fatigue for patients with EBV infection.

Correlation Analyses

For each of the 5 symptom domains, Pearson correlations between gene expression levels and principal components analysis-derived indices were calculated. Correlation tests were performed and *P* values (unadjusted and Benjamini-Hochberg adjusted) were obtained. All subjects and time points were included. Overabundance analyses were then used to compare the number of genes determined to have a correlation significantly different from 0, using a 5% (unadjusted) significance level, where the number expected by chance was obtained by randomly permuting the symptom scores 1000 times. Overabundance was assessed separately for each symptom domain and each infection.

Selection of Genes for Confirmatory PCR

The criteria for selection of genes for confirmation reflected the primary aim of identifying genes that were consistently differentially expressed across all 3 infective cohorts. Because the various analysis methods potentially captured associations with different aspects of the fatigue-associated symptom complex, genes identified by each method were included. For the paired, 2-sample, and GEE analyses, intersections of the top 2000 ranked genes for the 3 infections formed the basis for selection. For the correlation analysis, where no overabundance was observed for RRV infection, selection was made from the top-ranked genes significant at 5% for both EBV infection and Q fever for the overabundant severity symptom domain.

Analysis of Quantitative PCR Data

Correlation tests were performed to assess how well the gene expression levels measured by microarray correlated with quantitative PCR (qPCR) values. In addition, the paired, 2-sample, GEE, and correlation analyses were rerun using the correlated qPCR data set.

RESULTS

The mean age of the postinfective fatigue syndrome case subjects was 40 years (standard deviation [SD], 18 years) and that of the control group was 39 years (SD, 16 years; $P > .1$). There were 11 male subjects and 7 female subjects in both groups (Table 1). The initial sampling point was within 6 weeks of illness onset (mean, 4.2 weeks; SD, 1.8 weeks). Subjects in the EBV group were generally younger in age, which is typical of infectious mononucleosis, whereas subjects in the Q fever group were predominantly male and older, reflecting the occupational acquisition of this illness. The RRV group was heterogeneous in age and sex, consistent with the mosquito-borne transmission of this infection. These demographic characteristics are representative of the larger DIOS cohort [2].

Paired Analysis

There were 23 genes ranked in the top 2000 for all 3 cohorts that were consistently up- or down-regulated (Supplementary Table 1), although the fold changes observed were modest, ranging from ~0.6-fold (down-regulated in the samples from the early sick time point) to 2.3-fold (up-regulated in the samples from the early sick time point). There were no genes for which the adjusted significance level was <.05.

Two-sample Analysis

There were 15 genes that exhibited consistent up- or down-regulation (Supplementary Table 2). Again, estimated fold changes were modest, ranging from 0.6-fold to 2.4-fold. There were no genes for which the adjusted significance level was <.05.

Longitudinal Analysis

The longitudinal model identified 58 genes for which the estimated coefficients were required to have the same sign (Supplementary Table 3). Some very large effect sizes were evident, with coefficient estimates (log-odds ratios) ranging from −18.6 to 19.1. There were 13 genes with adjusted significance values of <.05, of which only 1 (cytochrome b-245, alpha polypeptide [*CYBA*]) was associated with fatigue in 2 of the 3 infective cohorts (EBV and Q fever).

Correlation Analysis

The overabundance analysis identified overall illness severity, as well as fatigue and neurocognitive disturbance, as being correlated for the EBV and Q fever cohorts, but not the RRV cohort ($P = .016$ and $P = .044$ for severity, $P = .017$ and $P = .038$ for fatigue, and $P = .019$ and $P = .023$ for neurocognitive disturbance for EBV and Q fever, respectively). Correlation tests identified 96 genes significant at 5% (unadjusted) for both EBV and Q fever for severity, 93 genes for the fatigue symptom domain, and 106 genes for neurocognitive disturbance (158 of the 223 genes were significant for 1 symptom domain, 58 for 2 symptom domains, and

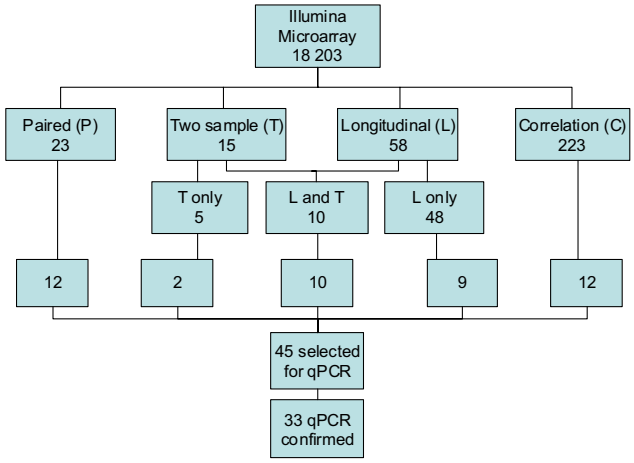

Figure 1. Selection of genes for confirmatory quantitative polymerase chain reaction (qPCR). For the 18 203 genes represented on the arrays (*first row*), 4 statistical analysis methods were applied (see the Methods) (*second row*). From the paired (P), 2-sample (T), and longitudinal (L) analyses, genes ranked in the top 2000 for all 3 infections were selected, giving 23, 15, and 58 genes, respectively. From the correlation (C) analyses, genes significant at 5% (unadjusted) for both Epstein-Barr virus (EBV) and Q fever, for the overabundant symptom domains of severity, fatigue, and neurocognitive disturbance, were selected, giving 223 genes. Ten of the selected genes were common to the 2-sample and longitudinal analyses (*third row*). A further selection was then made (*fourth row*). For the paired, 2-sample, and longitudinal analyses, the top 12, 12, and 19 genes, respectively, based on mean rank across the 3 infections, were selected. Ten genes were common to the 2-sample and longitudinal analyses. For the correlation analysis, the top 12 genes by mean rank for the 2 infections EBV and Q fever were selected. This process yielded 45 genes for confirmatory qPCR (*fifth row*), of which 33 (*sixth row*) were found to have significant correlations between qPCR values and normalized gene expression levels from the array analysis ($P < .1$ for correlactions).

7 for all 3 symptom domains) (Supplementary Table 4). The strength of the correlations ranged from $r = -0.61$ to $r = 0.63$. Three of the genes associated with neurocognitive disturbance in the EBV group had an adjusted significance level of <.05, including glutamate decarboxylase-2 (*GAD2*), sialoadhesion-1 (CD169; *SIGLEC-1*), and serine/threonine/tyrosine interacting protein (*STYX*). None of these genes was significantly correlated (adjusted *P* value) with neurocognitive disturbance in the Q fever group.

Selection of Genes for Confirmatory qPCR

For the paired, 2-sample, and longitudinal analyses, the first-level selection included genes appearing in the top 2000 rank lists for all 3 infections (Figure 1). At the second level, a further selection was made on the basis of the mean rank across the 3 infections, with the top 12 genes by mean rank chosen for the paired and 2-sample analyses, and the top 19 genes by mean rank chosen for the longitudinal analysis. Ten genes were in common for the 2-sample and longitudinal analyses. For the correlation analysis, the top 12 genes by mean rank for the 2 infections EBV and Q fever were chosen from the genes that were significant at 5%.

Analysis of qPCR Data

Of the 45 genes selected for confirmatory qPCR, 33 showed a correlation at the 10% level with gene expression levels determined by microarray (Table 2). The paired, 2-sample, and longitudinal analyses were then repeated using these 33 genes, to seek those with significant differences in a consistent direction for at least 2 of the 3 infections. In the repeated correlation analysis, no genes were found to be significantly correlated for both EBV infection and Q fever in association with severity, fatigue, or neurocognitive disturbance at a 5% significance level. With a more liberal significance threshold of 10% (unadjusted), 6 genes were significant: actin, alpha2, smooth muscle, aorta (*ACTA2;* up-regulated in postinfective fatigue syndrome in the paired analysis for EBV [$P = .084$] and RRV [$P = .020$]); junction plakoglobin (*JUP;* down-regulated in postinfective fatigue syndrome in the paired analysis for EBV [$P = .063$] and Q fever [$P = .092$]); Fanconi anemia, complementation group E (*FANCE;* up-regulated in postinfective fatigue syndrome in the longitudinal analysis for EBV [$P = .025$; GEE coefficient, 0.649] and down-regulated for RRV [$P = .038$; GEE coefficient, -0.723]); solute carrier family 27 member 6, fatty acid transporter (*SLC27A6;* up-regulated in postinfective fatigue syndrome in the longitudinal analysis for RRV [$P = .058$; GEE coefficient, 0.089] and down-regulated for Q fever [$P = .040$; GEE coefficient, -0.171]); teashirt zinc finger homeobox 2 (*TSHZ2;* correlated with fatigue for EBV [$P = .022$] and Q fever [$P = .084$]); and lectin, galactose-binding soluble 3, binding protein (*LGALS3BP;* correlated with neurocognitive disturbance for EBV [$P = .009$] and Q fever [$P = .077$]).

Table 2. Correlation Tests for Polymerase Chain Reaction ΔCt Values With Normalized Gene Expression

Gene symbol	Correlation	*P*	Confirmation[a]
ABCC5	−0.4746	<.0001	Yes
ACTA2	−0.8012	<.0001	Yes
ALG8	−0.5737	<.0001	Yes
APBA2	−0.3152	.0005	Yes
ATP5J	−0.1670	.0706	Yes
C10orf4	−0.4718	<.0001	Yes
C1orf35	−0.1090	.2399	No
C20orf72	−0.6459	<.0001	Yes
CBX3	0.1233	.1834	No
CCNB2	−0.6285	<.0001	Yes
COX6B1	−0.2444	.0076	Yes
CRB2	0.1300	.3535	No
CSDE1	−0.1463	.1140	No
CYBASC3	−0.3823	<.0001	Yes
EEF1B2	0.1347	.1460	No
EMP3	−0.2759	.0025	Yes
FANCE	−0.5563	<.0001	Yes
FCER1G	−0.5324	<.0001	Yes
HSPBP1	−0.0734	.4293	No
JUP	−0.7702	<.0001	Yes
LEPROTL1	−0.5851	<.0001	Yes
LGALS3BP	−0.3085	.0007	Yes
LMAN1	0.1807	.0502	Yes
LTF	−0.5998	<.0001	Yes
MAD1L1	0.0350	.7067	No
MRPL20	−0.3566	.0001	Yes
NDUFA1	−0.2832	.0019	Yes
NEDD8	−0.1019	.2722	No
NEFH	−0.1078	.2905	No
PFN1	−0.2245	.0145	Yes
PNPO	−0.5085	<.0001	Yes
PPM1G	0.2031	.0274	Yes
PRR11	−0.5680	<.0001	Yes
PVRIG	−0.3680	<.0001	Yes
RAB8A	−0.3172	.0005	Yes
RAG1AP1	−0.0194	.8352	No
RPP40	−0.5632	<.0001	Yes
SCRN1	−0.5483	<.0001	Yes
SERPINB6	−0.3599	.0001	Yes
SLC27A6	−0.2841	.0048	Yes
SLC35B3	−0.7827	<.0001	Yes
SSR2	−0.0917	.3236	No
TAX1BP1	−0.2180	.0177	Yes
TSHZ2	−0.6572	<.0001	Yes
YDJC	−0.0776	.4037	No

Abbreviations: Ct, cycle threshold; orf, open reading frame.

[a] $P < .10$ for correlation test.

Comparison of Genes of Interest With Previous Studies

A review of differentially expressed genes in published case-control studies of CFS or postinfective fatigue syndrome revealed only 6 genes that have been identified by >1 independent

Table 3. Genes Associated With Chronic Fatigue Syndrome in More Than One Case-Control Study

Gene symbol	Gene name	References
CXCR4	C-X-C chemokine receptor type 4	[14, 15, 28]
LTF	Lactotransferrin	[28], present study[a]
SH2D1B	SH2 domain containing 1B	[29, 48][b]
BCOR	BCL6 co-repressor	[15, 26]
BRMS1	Breast cancer metastasis suppressor 1	[15], present study[a]
ENSA	Endosulfine alpha	[49], present study[a]
DCTN1	Dynactin, dynein activator complex	[11, 30]
ARF1	ADP-ribosylation factor	[11, 30]
CEACAM	Carcinoembryonic antigen-related cell adhesion molecule family	[30, 50]
EGR1	Early growth response 1	[14, 30]
PRKAR1A	Protein kinase A regulatory subunit 1A	[14, 30]
Defensin	Defensin family	[20, 30]

[a] Identified in the microarray analysis only, not confirmed by quantitative polymerase chain reaction.

[b] It is assumed for this purpose that Gulf War illness is synonymous with chronic fatigue syndrome.

group (summarized in Table 3). None were shared with the genes listed above.

DISCUSSION

To our knowledge, this is the first report of a longitudinal analysis of the peripheral blood transcriptome in patients with postinfective fatigue syndrome that incorporates multiple comparators to minimize the likelihood of false discovery, including within-subject (sick versus well), between-subject (postinfective fatigue syndrome cases versus recovered controls), and between-cohort (EBV versus Q fever versus RRV) analyses. In addition, the data set described here includes empirical exploration of the potential gene expression correlates of the syndrome as a whole (ie, caseness for postinfective fatigue) and the major symptom domains considered in isolation (ie, fatigue, pain. mood disturbance, and neurocognitive disturbance). The major finding reported here with the use of this comprehensive approach was that no genes were consistently associated with the illness.

A key premise in these analyses was that the cardinal manifestations of the postinfective fatigue syndrome are stereotyped, and hence that gene expression correlates should also be shared, regardless of the original infective trigger. We previously reported the results of a planned contrasts analysis of the patterns of resolution of the major symptom domains of the postinfective fatigue syndrome in the DIOS cohort, and we revealed that only pain, but not fatigue, mood disturbance, or neurocognitive disturbance, differed in prevalence and natural history in the postinfective fatigue syndrome [37]. We also found that the incidence of CFS was closely comparable between the 3 infections [37]. Given these data, the finding here of the lack of a coherent set of gene expression correlates across cohorts argues against the validity of previously proposed signatures for postinfective fatigue syndrome [38] or CFS [39].

One of the most striking findings was the relative lack of variance in gene expression levels within subjects over $\geq$12 months, with differences consistently $<$2.3-fold identified. It should be noted in this regard that although all subjects were recruited within 6 weeks of onset of illness, all were well beyond the acute febrile phase of the infection at the time of enrollment. Accordingly, the gene expression data set at the initial time point reflects the early postinfective period, rather than being in the midst of the acute sickness response. Hence, enhanced expression of genes such as pro-inflammatory cytokines, which feature in acute infection, was not evident [40, 41]. By contrast, this finding argues that 2 of the key methodological issues of concern in microarray studies were not present in the data set described here: sampling and technical biases arising from variations in sample collection, handling, and processing; and genetic mismatching in comparative analyses. The former relates to issues such as variations in the timing of blood sample collection linked to circadian influences on gene expression [42]; appropriate ex vivo handling of the peripheral blood samples to minimize the possibility of endotoxin contamination and leukocyte activation [43]; and run-to-run variation in the microarray hybridizations [44]. The bias associated with variations in genetic background between case and control subjects relates to the fact that many human genes are under strong genetic control mediated by commonly inherited variants [45]. The within-subject analysis reported here resolves most of this potential bias, and the negative findings are consistent with the previous report of negative results in a study of monozygotic twin pairs [14]. We cannot exclude the possibility of variations in gene expression associated with ex vivo handling of samples, including separation and storage of PBMCs, and it also remains plausible that alterations in gene expression of limited extent, or alterations within leukocyte subpopulations, may be found in CFS; however, this appears unlikely. Finally, because CFS is likely to be a heterogeneous illness [15–17], the findings here in relation to the postinfective fatigue syndrome subset may not be generalizable to all patients with CFS.

A growing body of evidence points to the central nervous system as the likely site of the pathophysiological disturbance underpinning CFS [5, 6, 46]. The postinfective fatigue syndrome illness is likely to be truly postinfective—that is, unassociated with ongoing active replication of the triggering agent. The likely negative resolution of the putative association between xenotropic murine leukemia virus-related virus (XMRV) is consistent with this notion [47]. The data presented here suggest that

further investigation of the peripheral blood transcriptome is not warranted.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://www.oxfordjournals.org/our_journals/jid/).

Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. All authors contributed substantially to the conception and design, analysis and interpretation of data, drafting of the manuscript, critical revision, and final approval of the manuscript.

Financial support. The Dubbo Infection Outcomes Study was supported by the National Health and Medical Research Council of Australia (project grants 157092 and 157062; and Practitioner Fellowship 510246 to A. L.); and the US Centers for Disease Control and Prevention (Cooperative Research Agreement grant U50/CCU019851-01).

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Candy B, Chalder T, Cleare AJ, et al. Predictors of fatigue following the onset of infectious mononucleosis. Psychol Med **2003**; 33:847–55.
2. Cameron B, Bharadwaj M, Burrows J, et al. Prolonged illness after infectious mononucleosis is associated with altered immunity but not with increased viral load. J Infect Dis **2006**; 193:664–71.
3. White PD. What causes prolonged fatigue after infectious mononucleosis: and does it tell us anything about chronic fatigue syndrome? J Infect Dis **2007**; 196:4–5.
4. Petersen I, Thomas JM, Hamilton WT, White PD. Risk and predictors of fatigue after infectious mononucleosis in a large primary-care cohort. QJM **2006**; 99:49–55.
5. Afari N, Buchwald D. Chronic fatigue syndrome: a review. Am J Psychiatry **2003**; 160:221–36.
6. Prins JB, van der Meer JWM, Bleijenberg G. Chronic fatigue syndrome. Lancet **2006**; 367:346–55.
7. Lyall M, Peakman M, Wessely S. A systematic review and critical evaluation of the immunology of chronic fatigue syndrome. J Psychosom Res **2003**; 55:79–90.
8. Cameron B, Hirschberg DL, Rosenberg-Hassan Y, Ablashi D, Lloyd AR. Serum cytokine levels in postinfective fatigue syndrome. Clin Infect Dis **2010**; 50:278–9.
9. Vernon SD, Unger ER, Dimulescu IM, Rajeevan M, Reeves WC. Utility of the blood for gene expression profiling and biomarker discovery in chronic fatigue syndrome. Dis Markers **2002**; 18:193–9.
10. Powell R, Ren J, Lewith G, Barclay W, Holgate S, Almond J. Identification of novel expressed sequences, up-regulated in the leucocytes of chronic fatigue syndrome patients. Clin Exp Allergy **2003**; 33:1450–6.
11. Grans H, Nilsson P, Evengard B. Gene expression profiling in the chronic fatigue syndrome. J Intern Med **2005**; 258:388–90.
12. Kerr JR, Petty R, Burke B, et al. Gene expression subtypes in patients with chronic fatigue syndrome/myalgic encephalomyelitis. J Infect Dis **2008**; 197:1171–84.
13. Zhang L, Gough J, Christmas D, et al. Microbial infections in eight genomic subtypes of chronic fatigue syndrome/myalgic encephalomyelitis. J Clin Pathol **2010**; 63:156–64.
14. Byrnes A, Jacks A, Dahlman-Wright K, et al. Gene expression in peripheral blood leukocytes in monozygotic twins discordant for chronic fatigue: no evidence of a biomarker. PLoS One **2009**; 4: e5805.
15. Hickie I, Lloyd A, Hadzi-Pavlovic D, Parker G, Bird K, Wakefield D. Can the chronic fatigue syndrome be defined by distinct clinical features? Psychol Med **1995**; 25:925–35.
16. Wilson A, Hickie I, Hadzi-Pavlovic D, et al. What is chronic fatigue syndrome? Heterogeneity within an international multicentre study. Aust N Z J Psychiatry **2001**; 35:520–7.
17. Vollmer-Conna U, Aslakson E, White PD. An empirical delineation of the heterogeneity of chronic unexplained fatigue in women. Pharmacogenomics **2006**; 7:355–64.
18. Fang H, Xie Q, Boneva R, Fostel J, Perkins R, Tong W. Gene expression profile exploration of a large dataset on chronic fatigue syndrome. Pharmacogenomics **2006**; 7:429–40.
19. Aslakson E, Vollmer-Conna U, White PD. The validity of an empirical delineation of heterogeneity in chronic unexplained fatigue. Pharmacogenomics **2006**; 7:365–73.
20. Carmel L, Efroni S, White PD, Aslakson E, Vollmer-Conna U, Rajeevan MS. Gene expression profile of empirically delineated classes of unexplained chronic fatigue. Pharmacogenomics **2006**; 7:375–86.
21. Vernon SD, Reeves WC. The challenge of integrating disparate high-content data: epidemiological, clinical and laboratory data collected during an in-hospital study of chronic fatigue syndrome. Pharmacogenomics **2006**; 7:345–54.
22. Whistler T, Taylor R, Craddock RC, Broderick G, Klimas N, Unger ER. Gene expression correlates of unexplained fatigue. Pharmacogenomics **2006**; 7:395–405.
23. Broderick G, Craddock RC, Whistler T, Taylor R, Klimas N, Unger ER. Identifying illness parameters in fatiguing syndromes using classical projection methods. Pharmacogenomics **2006**; 7:407–19.
24. Kerr JR, Burke B, Petty R, et al. Seven genomic subtypes of chronic fatigue syndrome/myalgic encephalomyelitis: a detailed analysis of gene networks and clinical phenotypes. J Clin Pathol **2008**; 61:730–9.
25. Kerr JR. Gene profiling of patients with chronic fatigue syndrome/myalgic encephalomyelitis. Curr Rheumatol Rep **2008**; 10:482–91.
26. Fostel J, Boneva R, Lloyd A. Exploration of the gene expression correlates of chronic unexplained fatigue using factor analysis. Pharmacogenomics **2006**; 7:441–54.
27. Cannon TD, Keller MC. Endophenotypes in the genetic analyses of mental disorders. Annu Rev Clin Psychol **2006**; 2:267–90.
28. Gow JW, Hagan S, Herzyk P, Cannon C, Behan PO, Chaudhuri A. A gene signature for post-infectious chronic fatigue syndrome. BMC Med Genomics **2009**; 2:38.
29. Cameron B, Galbraith S, Zhang Y, et al. Gene expression correlates of postinfective fatigue syndrome after infectious mononucleosis. J Infect Dis **2007**; 196:56–66.
30. Robertson P, Beynon S, Whybin R, et al. Measurement of EBV-IgG anti-VCA avidity aids the early and reliable diagnosis of primary EBV infection. J Med Virol **2003**; 70:617–23.
31. Hickie I, Davenport T, Hadzi-Pavlovic D, et al. Development of a simple screening tool for common mental disorders in general practice. Med J Aust **2001**; 175:S10–7.
32. Hadzi-Pavlovic D, Hickie I, Wilson A, Davenport T, Lloyd A, Wakefield D. Screening for prolonged fatigue syndromes: statistical and longitudinal validation of the SOFA scale. Soc Psychiatry Psychiatr Epidemiol **2000**; 35:471–9.
33. Koschera A, Hickie I, Hadzi-Pavlovic D, Wilson A, Lloyd A. Prolonged fatigue, anxiety and depression: exploring relationships in a primary care sample. Aust N Z J Psychiatry **1999**; 33:545–52.
34. Panelli M, Wang E, Phan G, et al. Gene expression profiling of the response of peripheral blood mononuclear cells and melanoma metastases to systemic IL-2 administration. Genome Biol **2002**; 3:1–17.
35. Smyth G. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **2004**; 3:article 3.

36. Hojsgaard S, Jun Y, Halekoh U. The R package geepack for generalized estimating equations. J Stat Softw **2005**; 15:1–11.
37. Hickie I, Davenport T, Wakefield D, et al. Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: prospective cohort study. BMJ **2006**; 333:575–8.
38. Gow J, Hagan S, Herzyk P, Cannon C, Behan P, Chaudhuri A. A gene signature for post-infectious chronic fatigue syndrome. BMC Med Genomics **2009**; 2:38.
39. Kerr JR, Christian P, Hodgetts A, et al. Current research priorities in chronic fatigue syndrome/myalgic encephalomyelitis: disease mechanisms, a diagnostic test and specific treatments. J Clin Pathol **2007**; 60: 113–6.
40. Ramilo O, Allman W, Chung W, et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. Blood **2007**; 109:2066–77.
41. Zilliox MJ, Moss WJ, Griffin DE. Gene expression changes in peripheral blood mononuclear cells during measles virus infection. Clin Vaccine Immunol **2007**; 14:918–23.
42. Merrow M, Spoelstra K, Roenneberg T. The circadian cycle: daily rhythms from behaviour to genes. EMBO Rep **2005**; 6:930–5.
43. Riches P, Gooding R, Millar BC, Rowbottom AW. Influence of collection and separation of blood samples on plasma IL-1, IL-6 and TNF-alpha concentrations. J Immunol Methods **1992**; 153: 125–31.
44. MAQC Consortium. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol **2006**; 24:1151–61.
45. Whitney A, Diehn M, Popper SJ, et al. Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci U S A **2003**; 100:1896–901.
46. Cho HJ, Skowera A, Cleare A, Wessely S. Chronic fatigue syndrome: an update focusing on phenomenology and pathophysiology. Curr Opin Psychiatry **2006**; 19:67–73.
47. Hue S, Gray ER, Gall A, et al. Disease-associated XMRV sequences are consistent with laboratory contamination. Retrovirology **2010**; 7:111.
48. Whistler T, Fletcher M, Lonergan W, et al. Impaired immune function in Gulf War illness. BMC Med Genomics **2009**; 2:12.
49. Vernon S, Whistler T, Cameron B, Hickie I, Reeves W, Lloyd A. Preliminary evidence of mitochondrial dysfunction associated with post-infective fatigue after acute infection with Epstein Barr virus. BMC Infect Dis **2006**; 6:15.
50. Kaushik N, Fear D, Richards SCM, et al. Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome. J Clin Pathol **2005**; 58:826–32.