MAJOR ARTICLE

Gene Expression Correlates of Postinfective Fatigue Syndrome after Infectious Mononucleosis

Barbara Cameron,[1] Sally Galbraith,[2] Yun Zhang,[2] Tracey Davenport,[4] Ute Vollmer-Conna,[3] Denis Wakefield,[1] Ian Hickie,[4] William Dunsmuir,[2] Toni Whistler,[5] Suzanne Vernon,[5] William C. Reeves,[5] and Andrew R. Lloyd,[1] for the Dubbo Infection Outcomes Study

[1]School of Medical Sciences, [2]School of Mathematics, and [3]School of Psychiatry, University of New South Wales, and [4]Brain and Mind Research Institute, Sydney University, Sydney, Australia; [5]Division of Viral and Rickettsial Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia

(See the editorial commentary by White, on pages 4–5.)

Background. Infectious mononucleosis (IM) commonly triggers a protracted postinfective fatigue syndrome (PIFS) of unknown pathogenesis.

Methods. Seven subjects with PIFS with 6 or more months of disabling symptoms and 8 matched control subjects who had recovered promptly from documented IM were studied. The expression of 30,000 genes was examined in the peripheral blood by microarray analysis in 65 longitudinally collected samples. Gene expression patterns associated with PIFS were sought by correlation with symptom factor scores.

Results. Differential expression of 733 genes was identified when samples collected early during the illness and at the late (recovered) time point were compared. Of these genes, 234 were found to be significantly correlated with the reported severity of the fatigue symptom factor, and 180 were found to be correlated with the musculoskeletal pain symptom factor. Validation by analysis of the longitudinal expression pattern revealed 35 genes for which changes in expression were consistent with the illness course. These genes included several that are involved in signal transduction pathways, metal ion binding, and ion channel activity.

Conclusions. Gene expression correlates of the cardinal symptoms of PIFS after IM have been identified. Further studies of these gene products may help to elucidate the pathogenesis of PIFS.

In industrialized countries, 40%–65% of primary Epstein-Barr virus (EBV) infections occur asymptomatically during early childhood [1, 2]. In contrast, primary EBV infection in young adults often causes symptomatic infectious mononucleosis (IM). Most cases of acute IM resolve within several weeks without sequelae, but some individuals experience a prolonged and disabling illness marked by fatigue extending over weeks or months. Prospective cohort studies examining the kinetics of recovery from acute IM [3–5] have revealed that almost 50% of subjects had ongoing symptoms at 2 months after onset and that ~10% had disabling symptoms marked by fatigue lasting 6 months or more. These subjects did not have clinical features of chronic, active EBV (CAEBV) infection, which is attributable to congenital [6] or acquired [7, 8] impairments of T cell immunity. Similarly, detailed medical and psychiatric assessments conducted in the Dubbo Infection Outcomes Study (DIOS) [5] did not reveal an alternative medical or psychiatric explanation for this postinfective fatigue syndrome (PIFS), indicating that subjects with PIFS represent a subset of the more heterogeneous and enigmatic clinical disorder termed "chronic fatigue syndrome" (CFS) [9, 10].

We recently reported the outcomes of a detailed assessment of virological and immunological correlates of PIFS in a case-control series of DIOS subjects followed from the onset of acute IM [11]. There was no difference in cellular EBV load at any time point between the case subjects, who developed PIFS, and the control subjects, who recovered promptly. Although minor alterations in the kinetics of both antibody and

Received 5 December 2006; accepted 12 January 2007; electronically published 24 May 2007.

Potential conflicts of interest: none reported.

Financial support: National Health and Medical Research Council of Australia (project grants 157092 and 157062); US Centers for Disease Control and Prevention (cooperative research agreement U50/CCU019851-01).

Correspondence: Prof. Andrew Lloyd, Centre for Infection and Inflammation Research, School of Medical Sciences, University of New South Wales, Sydney 2052, Australia (A.Lloyd@unsw.edu.au).

The Journal of Infectious Diseases 2007; 196:56–66

0022-1899/2007/19601-0011$15.00
DOI: 10.1086/518614

Table 1. Subject groups, symptom scores, and sampling time points (T1–T4) for the microarray analysis.

Subject[a] (sex, age in years)	Symptom scores[b] by time after onset								
	T1		T2		T3		T4		
	0–3 weeks	3–6 weeks	6–9 weeks	9–12 weeks	3–6 months	6–9 months	9–12 months	>12 months	>12 months
PIFS1 (F, 49)		8	7	8	8	8		4	5
PIFS2 (F, 17)		3	6		9	10	11	10	5
PIFS3 (F, 23)		11	8				10	6	
PIFS4 (F, 17)	6	9			4		7	0	
PIFS5 (M, 18)		11	NA	8	7		3	2	0
PIFS6 (M, 23)		4			5			0	
PIFS7 (F, 19)		5		8	3				
C1 (F, 34)		5	4					0	
C2 (M, 17)			8	0					
C3 (M, 19)		9	2			0		0	
C4 (M, 48)		3		NA		0		1	
C5 (M, 18)	3	1			0			0	
C6 (M, 18)	0	2	0					1	
C7 (F, 19)		2	2	1				0	
C8 (F, 18)	1	0	0					0	

NOTE. Shaded and unshaded areas indicate the illness and the recovery period, respectively, for each subject. NA, not available.

[a] Case patients with postinfective fatigue syndrome are indicated by "PIFS"; matched control subjects who recovered within 6 weeks of onset are indicated by "C."

[b] Symptom score on the SOMA subscale of the Somatic and Psychological Health Report (possible range, 0–12; a score of 3 or more indicates a clinically significant fatigue state).

T cell responses to EBV were evident, these did not correlate with the timing of recovery, arguing against the popular immunological and virological hypotheses of the pathogenesis of PIFS [12, 13]. In combination with available evidence from other studies of patients with CFS, these data point to the central nervous system (CNS) as the likely site of the pathophysiological disturbance [13–16]. Thus, we predicted that neurochemical and neuroinflammatory genes would be differentially expressed in the peripheral blood of subjects with PIFS.

Accordingly, the present study adopted a gene discovery approach as a novel strategy to elucidate the pathophysiology of PIFS. Peripheral blood was chosen for study, first because it was readily available; second because the utility of peripheral blood gene expression to explore the pathogenesis of complex diseases of the CNS has recently been demonstrated in predicting emergent posttraumatic stress disorder [17] and in distinguishing subjects with schizophrenia or bipolar disorder from healthy control subjects [18]; and finally because preliminary studies examining samples collected from a group of subjects during acute IM discovered novel gene expression correlates of the symptoms of the acute sickness response [19]. Thus, the present study used microarray technology to examine gene expression in a matched case-control series of subjects followed from shortly after the onset of acute IM that was of short duration or that persisted into PIFS.

SUBJECTS, MATERIALS, AND METHODS

Subjects. Participants were enrolled in DIOS after presentation with symptoms of acute IM and detection of IgM antibodies against EBV capsid antigen. Follow-up was conducted at regular intervals for 12 months or more. Provisional serological diagnoses were confirmed by testing longitudinally collected serum samples [20].

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 a self-report questionnaire, the Somatic and Psychological Health Report [21, 22]. A score of 3 or more (of a possible 12) on a validated subscale (called "the SOMA") was used to designate clinically significant fatigue states [23–25].

In those subjects with persistent symptoms beyond 3 months (designated as having provisional PIFS cases), structured medical and psychiatric assessments as well as laboratory investigations to exclude CAEBV infection or unrelated causes of illness were undertaken in accordance with the diagnostic criteria for CFS [9, 10]. Seven subjects with PIFS (i.e., those who had unexplained illness persisting for 6 months or more after onset of symptoms and met the diagnostic criteria for CFS) and 8 control subjects who had recovered more promptly,

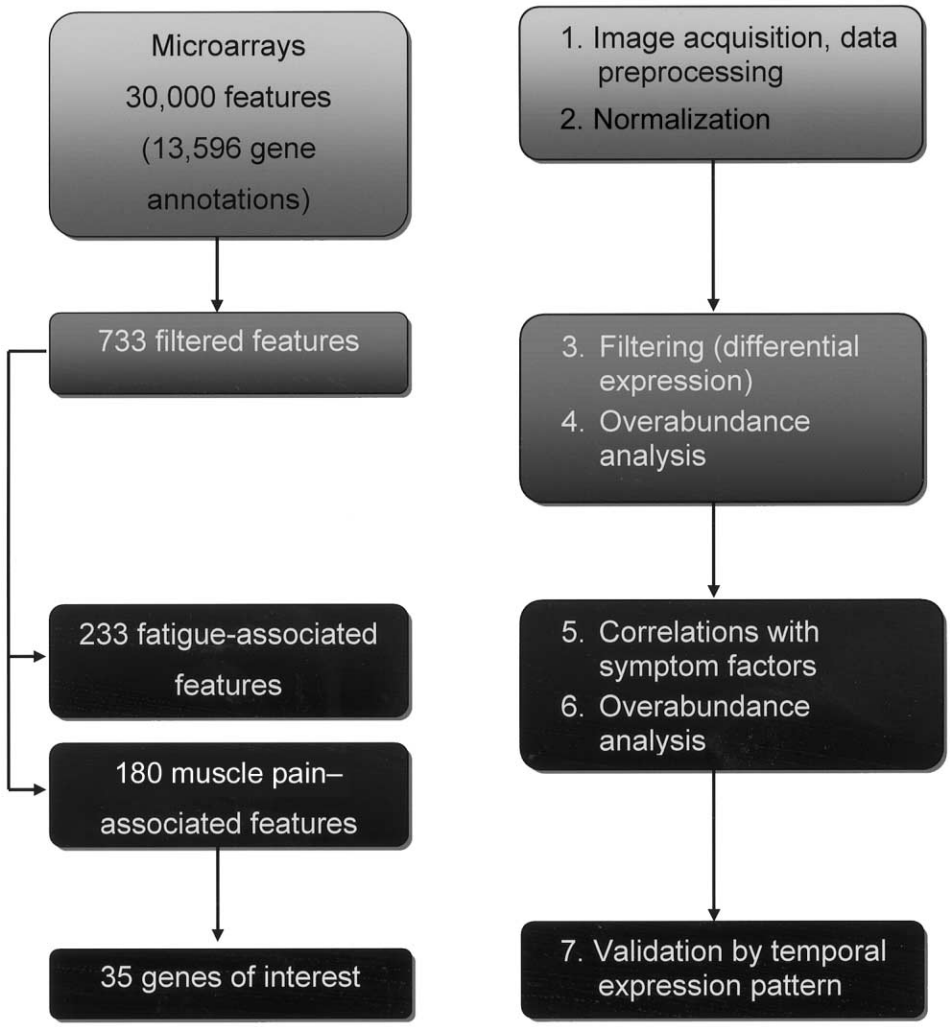

Figure 1. Schematic outline of the data analysis process

matched as a group by age and sex, were selected for the present study [5].

To allow investigation of the gene expression correlates of the symptom complex, scores for each subject at each time point for the 6 symptom factors (as described elsewhere [5])—fatigue, musculoskeletal pain, mood disturbance, neurocognitive disturbance, acute sickness, and irritability—were calculated from their self-report data sets. The study protocol was approved by the relevant institutional review boards. Written, informed consent was provided by all subjects.

Specimens and laboratory methods. Blood samples were collected in the morning and transported to the laboratory within 6 h. Then, peripheral blood mononuclear cells (PBMCs) were separated (Lymphoprep; AXIS-SHIELD) and cryopreserved with 10% DMSO (Sigma) and 50% autologous plasma, and aliquots were stored in the vapor phase of liquid nitrogen. Subsequently, the thawed PBMCs were lysed in Tri Reagent (Sigma). RNA was extracted and quantified by spectrophotometry, and quality was evaluated by denaturing gel electrophoresis.

Glass arrays (MWG Biotech) carrying 50mer oligonucleotides for 30,000 genes (10,000 on each of the 3 arrays, designated A, B, and C) were used. The A array bore 10,000 well-characterized genes; the B array carried a mix of known genes and expressed sequence tags (ESTs); and the C array bore all ESTs. Biotinylated cDNA probes were prepared from 1 μg of sample RNA as described elsewhere [26] and were hybridized to the arrays on the Ventana instrument (Ventana Medical Systems). Hybridization was for 8 h at 42°C with the ChipMap kit, with three 10-min stringency washes at 42°C (NaCl-Na citrate buffer at 2×, 1×, and 0.1×). This was followed by a 30-min incubation in streptavidin-labeled gold-particle solution (RLS system; Invitrogen [previously Genicon Sciences]) before vigorous washing to remove the oil, application of a liquid optical coating, and air drying.

A total of 65 samples were included, representing from 3 to

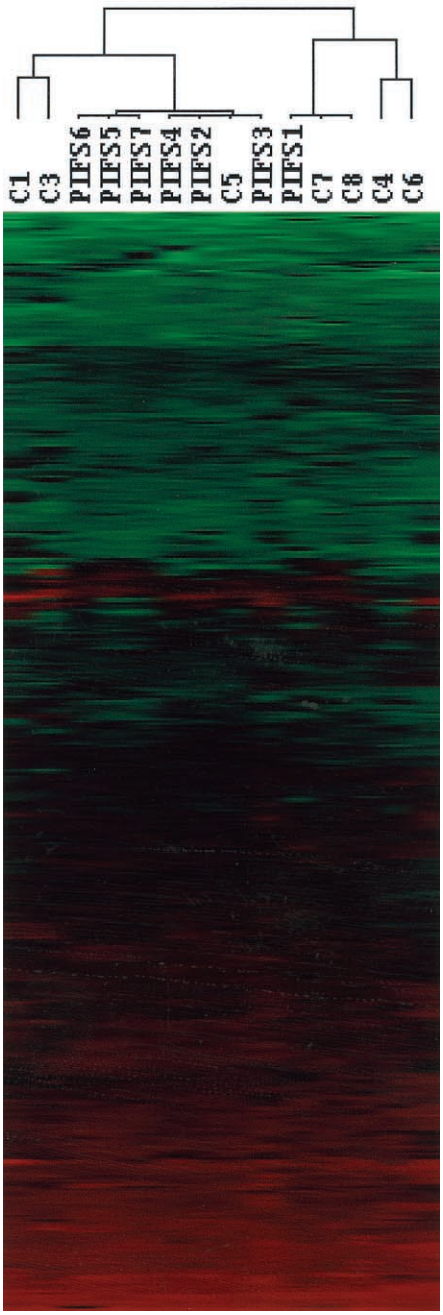

Figure 2. Cluster analysis of the filtered gene list including data from the baseline time point (T1). The gene expression data from T1 for the 733 features included in the filtered list were used in an unsupervised hierarchical cluster analysis, to identify a gene expression signature early during the course of infectious mononucleosis that predicted postinfective fatigue syndrome (PIFS) or recovery. Individual genes are in rows, and subjects ("PIFS" indicates case subjects; "C" indicates control subjects) are in columns.

7 time points per subject (table 1). RNA of sufficient yield and quality (a 28S:18S ratio of 1.8–2.0) was available to hybridize with all 3 arrays for 43 sampling points, whereas only hybridization with the A and B arrays was conducted for 13 samples, and, for a further 9 samples, hybridization with the A array alone was performed. All subjects had hybridizations preformed with all 3 arrays for at least 2 sampling points. 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 where possible, to control for run-to-run effects.

Data handling and analysis. Arrays were scanned (GSD-501; Invitrogen [previously Genicon Sciences]) with settings chosen to saturate a minimum of 1 feature on each array. Array images were analyzed (ArrayVision RLS; Imaging Research) to remove unacceptable features (i.e., "flags" due to dust or other technical artifacts) after manual confirmation as well as to quantify the relative expression of each feature in comparison to background. The raw intensity values ranged from 0 to 64,000. Flagged features as well as blanks and *Arabidopsis* controls were removed from the analysis [27].

Normalization. The data were normalized within each array by assuming that the intensity values plus an array-specific constant, after $\log_2$ transformation, followed a normal distribution, with zero values representing left-censored observations. Parameters of the distribution were estimated by maximum likelihood (S.G. and W.D., submitted manuscript). This approach transformed the data to normality, removed artificial array-specific effects, and recognized left censoring of intensity values at zero.

Filtering. It was assumed that the majority of the genes would not be differentially expressed and that including these noninformative genes might distort the clustering and correlation analyses [28]. In addition, it was assumed that the genes of interest in relation to PIFS would be differentially expressed when comparing the early illness phase with recovery, consistent with our recent evidence that all of the phenotypic characteristics of the PIFS illness are present from onset but resolve slowly [5]. The filtering procedure therefore compared expression levels in samples collected from subjects with a SOMA score of 3 or more at baseline (T1; representing data from samples collected during the early symptomatic phase of IM) with levels in subjects with a SOMA score <3 by 9 months (T4) and also during the preceding 3 months (representing data from samples collected well after recovery from IM and PIFS) (table 1). A feature was deemed to be differentially expressed if a 2-sample *t* test for equality of the means of normalized expression levels between the 2 groups (not assuming equality of variances) resulted in a *P* value of .01 or less. Outlying data points in the second group (expression levels >1.5 times the interquartile range below the first quartile or above the third quartile) were excluded before performing the *t* test.

Overabundance analysis. To assess the significance of the correlations between symptom scores and expression data of the filtered set of features, overabundance analysis was per-

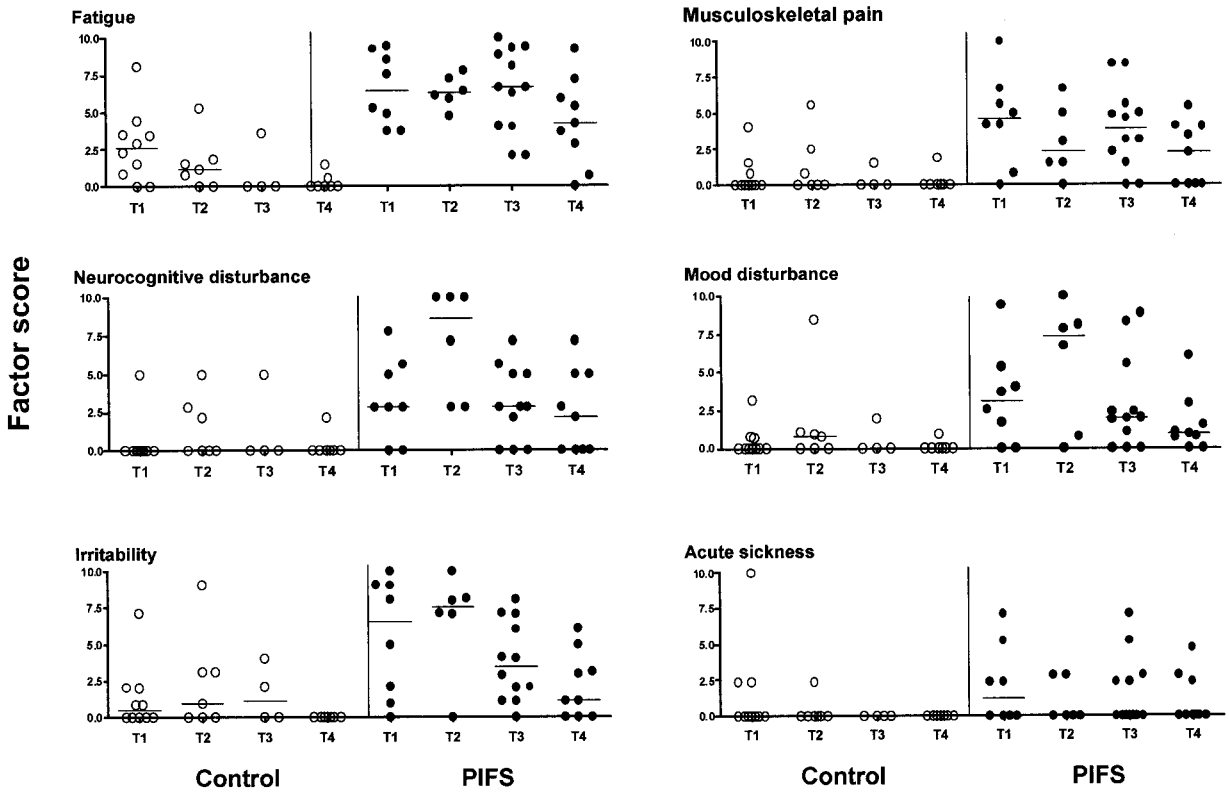

Figure 3. Symptom factor scores over time for case subjects with postinfective fatigue syndrome (PIFS) and control subjects. Normalized symptom factor scores for all data points for each subject at each categorized time point (T1–T4) were calculated [5]. Subjects with PIFS are represented by black circles, and control subjects are represented by white circles.

formed as described elsewhere [17]. This technique compared the number of features designated as being differentially expressed with the number expected by chance, which was obtained by randomly permuting the group labels 1000 times for different *P* value cutoffs. The 2-sample *t* test was performed for each permutation, and the number of *P* values less than each cutoff, averaged over all permutations, represented the expected number of features "differentially expressed" under the null hypothesis of no difference in expression levels between the 2 groups.

Clustering. Clustering was used to determine whether expression levels for the filtered features at baseline (T1) were able to classify subjects correctly according to case/control status. Clustering was performed using DoublePCluster software (available in the public domain ScoreGenes package [version December 2002]; see http://www.compbio.cs.huji.ac.il/scoregenes/), which implements an unsupervised hierarchical biclustering approach [17]. For subjects with >1 array at T1, the first array was used.

Correlation analyses. For each of the 6 symptom domains, Pearson correlations between symptom factor scores and expression levels were calculated for each feature. All subjects and time points were included in this analysis. To test the null hypothesis of zero correlation, an upper 1-sided *t* test was performed. Features with a *P* value <.05 were deemed to be significantly positively correlated, and those with a *P* value >.95 were deemed to be significantly negatively correlated. To assess the significance of the resulting sets of features, overabundance analyses were again performed using 1000 random permutations of the group labels.

Bioinformatics. The National Center for Biotechnology Information UniGene cluster ID as well as the RefSeq (reference sequence) and GenBank accession numbers were sought for each feature on the arrays by use of the SOURCE automated-annotation Web site (http://source.stanford.edu/). Of the original 30,000 features, 13,956 with gene annotations were identified. WebGestalt (WEB-based GEne SeT AnaLysis Toolkit) software (http://bioinfo.vanderbilt.edu/webgestalt/) [29], which includes GOTree Machine software, was used for comparative functional analysis. Gene ontology (GO) terms were sought for each annotated feature. For each GO term, the total number of features on the array belonging to that category was determined, before comparisons were made with the lists of symptom factor–correlated features identified in the analyses described above. Statistical analysis of the enrichment by GO category was completed using the hypergeometric test, which

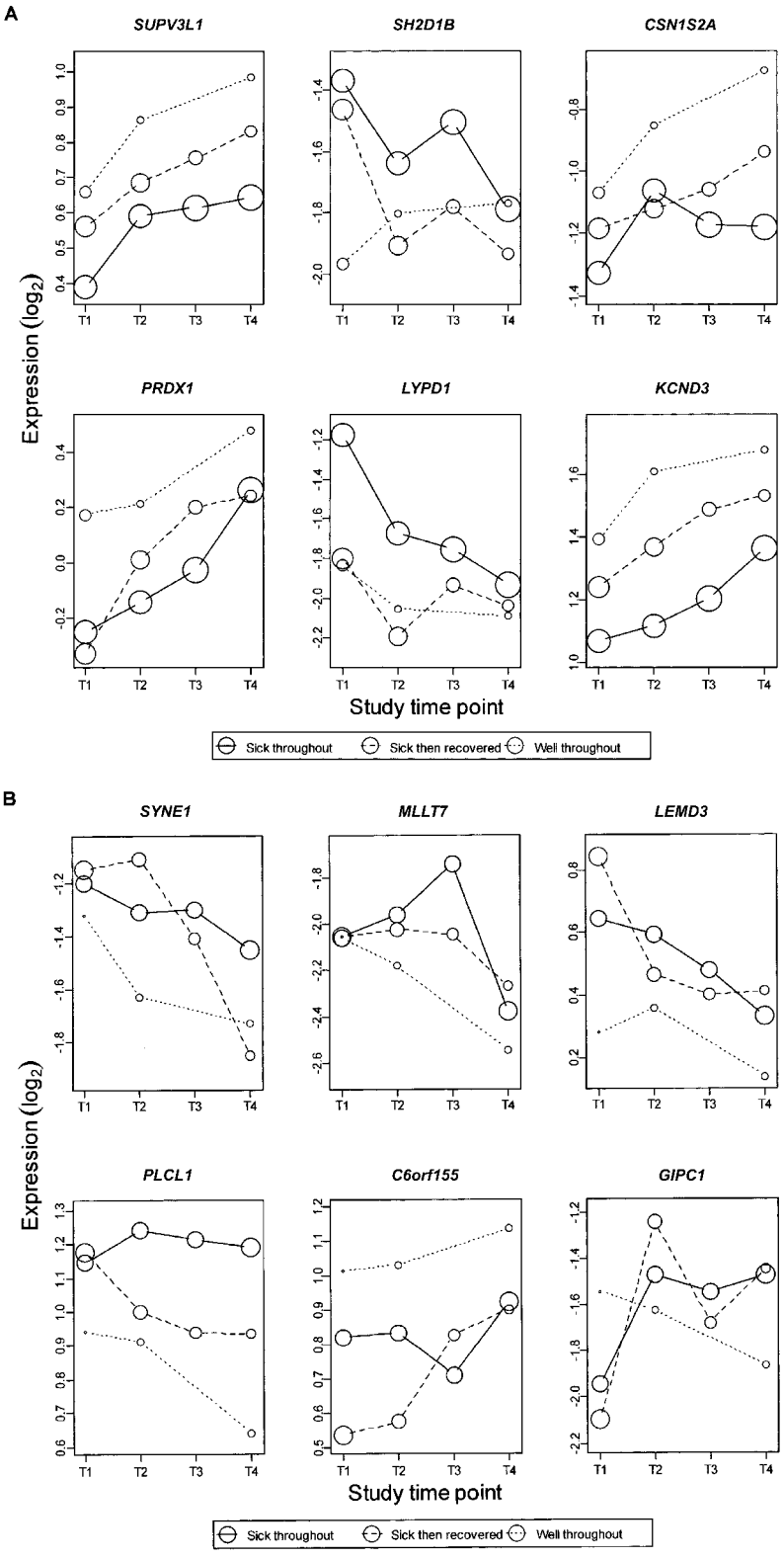

Figure 4. Expression of postinfective fatigue syndrome–associated genes over time in subjects with varied illness outcomes. Selected genes of interest of the 35 identified with temporal expression patterns consistent with the course of illness, associated with either fatigue *(A)* or musculoskeletal pain *(B),* are shown. The size of the symbols is proportional to the mean symptom factor score for the subgroup of subjects.

Table 2. Postinfective fatigue syndrome–associated genes: function, subcellular localization, tissue expression, and disease associations.

GenBank accession no.	Gene name	Gene symbol	Gene function	Subcellular localization	Tissue expression	Associated phenotype(s)
NM_003171	Suppressor of var 1, 3–like 1 *(S. cerevisiae)*	*SUPV3L1*	Cofactor of survivin in apoptosis suppression; ATP and RNA binding; helicase activity	Mitochondrion	Ubiquitous	
NM_182961	Spectrin repeat containing, nuclear envelope 1	*SYNE1*	Actin binding; laminin binding; Golgi and nuclear organization and biogenesis; myocyte differentiation	Golgi apparatus; cytoskeleton; nuclear envelope	Ubiquitous	
AC013478	Phospholipase C–like 1	*PLCL1*	Calcium ion binding; hydrolase activity; phospholipase C activity		Ubiquitous	
AL034421	Core-binding factor; runt domain, alpha subunit 2	*CBFA2T2*	Metal ion binding; regulation of transcription; cell proliferation	Nucleus	Ubiquitous	Translocation in acute myeloid leukemia produces a chimeric gene product associated with nuclear corepressor/histone deacetylase complex to block hematopoietic differentiation
AF013160	NADH dehydrogenase (ubiquinone) Fe-S protein 2, 49 kDa (NADH-coenzyme Q reductase)	*NDUFS2*	Metal ion binding; mitochondrial electron transport; NADH dehydrogenase (ubiquinone) activity; oxidative phosphorylation	Mitochondrial inner membrane	Ubiquitous	Mitochondrial complex I deficiency
NM_002574	Peroxiredoxin 1	*PRDX1*	Antioxidant enzyme activity; antiviral activity in CD8 T cells; cell proliferation; skeletal development	Cytoplasm	Ubiquitous	Increased expression in Alzheimer disease, Down syndrome, and lung injury
AL117330	CHK2 checkpoint homolog *(S. pombe)*	*CHEK2* (also *CDS1*)	ATP, Mg, nucleotide, and protein binding; serine/threonine kinase activity; response to DNA damage; cell cycle regulation	Nucleus	Ubiquitous	Associated with cancer susceptibility, including breast, colorectal, and prostate cancer
NM_005938	Myeloid/lymphoid or mixed-lineage leukemia	*MLLT7*	Transcription factor; cell cycle arrest; cell differentiation; negative regulation of angiogenesis, cell proliferation, and muscle cell differentiation; AKT and Ras signaling pathways	Nucleus; cytoplasm	Ubiquitous	
AC004908	Zinc finger protein 596	*ZNF596*	DNA and metal ion binding; DNA-dependent regulation of transcription	Nucleus	Ubiquitous	
AF351784	DnaJ (Hsp40) homolog, subfamily C, member 14	*DNAJC14*	Heat shock protein and unfolded protein binding; interacts with angiotensin receptor–1 (AGTR1), dopamine receptor–1 (DRD1), and lysosomal trafficking regulator (LYTR)	Membranes of the endoplasmic reticulum	Ubiquitous	
NM_016412	Insulin-like growth factor 2 antisense	*IGF2AS*	Growth factor inhibition		Brain; liver; placenta; plasma; pancreas	
NM_005505	Scavenger receptor class B, member 1	*SCARB1*	Apoptosis; cell adhesion; cholesterol metabolism	Plasma membrane	Ubiquitous	HCV entry cofactor; amyloid B protein interaction
AL365449	Sortilin-related VPS10 domain containing receptor 3	*SORCS3*	Neuropeptide signaling pathway	Membranes of the endosomes, Golgi, lysosomes, and nucleus	Brain; testis; cranial nerve; blood; liver; stomach; colon; muscle; larynx; tonsil; mammary gland	
NM_053282	SH2 domain containing 1B	*SH2D1B*	Intracellular signaling; NK cell–mediated cytotoxicity; interacts with NK cells; lymphocyte adhesion		Blood; spleen; thymus; kidney; stomach; skin; lung; muscle; testis	
NM_024882	Chromosome 6 open reading frame 155	*C6orf155*				
AK026814	Sorting nexin 25	*SNX25*	Signal transduction; phosphoinositide binding; protein binding		Ubiquitous	

AF255647	Transmembrane protein 163	*TMEM163*			Ubiquitous	
NM_003610	RAE1 RNA export 1 homolog *(S. pombe)*	*RAE1*	Interacts with NK cell lectins, nucleoporin; RNA binding and export; microtubule binding; mitotic spindle assembly	Cytoskeleton; nuclear membrane	Ubiquitous	
NM_005952	Metallothionein 1X	*MT1X*	Iron, zinc, copper, and cadmium ion binding; electron transport		Ubiquitous	
AC010974	LY6/PLAUR domain containing 1	*LYPD1*	GPI-anchored protein binding	Plasma membrane	Ubiquitous	
AC063956	Casein alpha s2–like A	*CSN1S2A*	Transporter activity	Extracellular region	Muscle	
AK002014	Chromosome 6 open reading frame 70	*C6orf70*			Ubiquitous	
NM_014319	LEM domain containing 3	*LEMD3*	Nuclear endoplasmic reticulum–associated degradation pathway; glycosylation of mammalian N-linked oligosaccharides; nucleotide binding	Nuclear inner membrane	Ubiquitous	Buschke-Ollendorff syndrome; melorheostosis with osteopoikilosis
NM_019893	N-acylsphingosine amidohydrolase (non–lysosomal ceramidase) 2	*ASAH2*	Ceramide, lipid, and sphingolipid metabolism signal transduction; apoptosis	Mitochondria; plasma membrane	Skin; bladder	
AF020762	Acyl-Coenzyme A binding domain containing 3	*ACBD3*	Maintenance of Golgi structure and function; hormonal regulation of steroid formation	Golgi membrane; cytoplasm; mitochondrion	Ubiquitous	
BC029816	Ovostatin 2	*OVOS2*	Endopeptidase inhibitor activity	Secreted protein	Testis; lung; eye; brain; lymph node; mammary gland; pituitary gland; bone; kidney; cranial nerve	
AL356315	Zinc finger and BTB domain containing 41	*ZBTB41*	Nucleic acid, protein, and zinc ion binding	Nucleus	Ubiquitous	
NM_014473	Dimethyladenosine transferase	*HSA9761*	rRNA modification and processing; transferase activity	Nucleus	Ubiquitous	
AF004813	Solute carrier family 4, sodium bicarbonate cotransporter, member 4	*SLC4A4*	Intracellular pH regulation; sodium ion binding and transport; anion exchange activity	Cell membrane	Ubiquitous	Proximal renal tubular acidosis with ocular abnormalities
NM_004512	Interleukin 11 receptor, alpha	*IL11RA*	Hematopoietin/interferon-class (D200-domain) cytokine receptor activity; Jak-STAT signaling pathway	Plasma membrane	Ubiquitous	High expression in Hodgkin lymphoma; possible association with multiple myeloma and prostate cancer
NM_004980	Potassium voltage-gated channel, Sha l-related subfamily, member 3	*KCND3*	Regulation of neurotransmitter release, heart rate, insulin secretion, neuronal excitability, epithelial electrolyte transport, smooth muscle contraction, and cell volume; metal ion, potassium, and protein binding	Plasma membrane; voltage-gated potassium channel complex	Brain; testis; prostate; lung; thyroid; mammary gland; colon; heart; adrenal gland	
NM_016598	Zinc finger, DHHC-type containing 3	*ZDHHC3*	Acyltransferase activity; metal ion binding	Vacuolar membrane	Ubiquitous	
NM_017594	GIPC PDZ domain containing family, member 1	*GIPC1*	Interacts with integrins, β-adrenergic receptor signaling pathway; spliceosomal assembly	Cytoplasm; plasma membrane; nucleus	Ubiquitous	
NM_017594	Small nuclear ribonucleoprotein polypeptide G	*SNRPG*	RNA and protein binding; RNA splicing; spliceosome assembly	Nucleus	Ubiquitous	
NM_002922	Regulator of G-protein signaling 1	*RGS1* (also *BL34*)	G-protein signaling; adenylate cyclase inhibition pathway; GTPase activator activity; calmodulin binding	Plasma membrane	Ubiquitous	

accounts for the problem of sampling without replacement associated with comparison of the filtered and symptom factor–correlated genes from the remaining features on the arrays. A GO category was considered to be differentially regulated if the significance level was <.01.

Finally, to validate the biological relevance of the symptom factor–correlated genes, the subject group was divided into 3 subgroups on the basis of the course of illness: those who remained symptomatic throughout the 9 months or more of follow-up ($n = 4$; subjects PIFS1–3 and PIFS7 in table 1); those who were symptomatic on enrollment but subsequently recovered ($n = 8$; subjects PIFS4–6 and C1–5); and those who had already recovered from IM shortly before enrollment in the cohort and remained symptom free over the period of prolonged follow-up ($n = 3$; subjects C6–8). For each gene, the mean normalized expression values and mean symptom factor scores for these 3 subject subgroups were plotted. Candidate genes were retained if (1) the pattern of change in expression over time was consistent with that predicted from the categorization of the subjects—that is, the mean intensity was highest in those who were symptomatic and lowest in those were not (or the converse for negatively correlated genes)—and (2) the pattern of recovery from illness over time was reflected by a 1.5-fold or greater ($\geqslant \log_2 0.59$) change in mean expression levels between the extremes of the data set. A single outlying data point in the mean trend lines was ignored, but the presence of 2 or more outlying data points led to the exclusion of that gene from further interest.

Functional and pathway information for the finalized list of genes was obtained from the BioCarta (http://www.biocarta.com/) and Kyoto Encyclopedia of Genes and Genomes (http://www.genome.ad.jp/kegg/kegg4.html) databases. Figure 1 provides a schematic summary of the complete data analysis process.

RESULTS

The subjects with PIFS included 2 males and 5 females with a mean age of 24 years (table 1). At enrollment, these subjects reported a mean of 22 days out of role and 14 days in bed since the onset of IM, whereas the control group, which included 5 males and 3 females with a mean age of 24 years, reported a mean of 17 days out of role and 9 days in bed. With the exception of 1 Hispanic individual, all subjects were white. All subjects had a clinical illness consistent with IM, featuring fever and pharyngitis. Generalized lymphadenopathy was evident in 10 subjects, splenomegaly in 1, and rash in 1. None of the group had preexisting medical illnesses that might have contributed to the symptom complex or influenced gene expression, with the possible exception of 1 case subject who had idiopathic epilepsy well controlled by sodium valproate therapy. Five subjects reported recent use of prescribed antibiotics (1 case and 4 control subjects) at baseline. Two females (1 case and 1 control subject) were taking the oral contraceptive pill. All subjects reported occasional use of simple analgesics, typically paracetamol, during the illness.

The cluster analysis sought a gene expression signature to distinguish case from control subjects during early illness (T1), to allow prediction of the subsequent development of PIFS. The solution dendrogram (figure 2) categorized the subjects into 2 broad groups, with 6 of the 7 case subjects in one arm and the remaining case subject (PIFS1) in the other arm. This subject was significantly older (49 years) than the other case subjects and was 1 of the 2 who had sustained illness over 12 months or more of follow-up. The 6 clustered PIFS cases were associated with 3 control subjects (C1, C3, and C5), who had no apparent distinguishing features. Cluster analysis of the T3 data set, which included the case subjects with 6 months or more of illness and the recovered control subjects, did not provide a coherent gene expression signature for PIFS.

Gene expression correlates of the 6 symptom domains were sought by analysis of the filtered gene list and the symptom factor scores for all subjects at all time points (figure 3). The fatigue factor was correlated, positively or negatively, with 197 genes, and the musculoskeletal pain factor was correlated with 138 genes. Overabundance analyses revealed that these 2, but not the other 4, symptom factors were associated significantly more commonly than by chance alone ($P < .0001$ for fatigue and $P = .007$ for musculoskeletal pain). Of these genes, 83 were associated with both factors, giving a combined list of 252 genes.

Of the 252 fatigue- and/or pain-associated genes, 35 were validated by analysis of the temporal course of the illness in relation to the gene expression pattern (figure 4). Analysis of the enrichment of these 35 genes by GO category did not identify recognized biological processes, molecular functions, or cellular components in which >1 gene from the disease-associated list was implicated, indicating significant enrichment in comparison to the GO categories associated with all annotated features on the arrays. Nevertheless, it is apparent that several members of the gene list are involved in similar biological themes, including signal transduction pathways, metal ion binding, and ion channel activity (table 2).

DISCUSSION

The present study provides the first comprehensive and longitudinal examination of the peripheral blood transcriptome in patients with well-characterized PIFS. Although peripheral blood is a complex tissue, a previous study revealed relatively restricted interindividual and within-individual variability in gene expression when studied by microarray analysis and also showed that this variance was markedly less than that observed in disease states [30]. In addition, recent data from the Microarray Quality Control project indicates good intraplatform

consistency across test sites as well as a high level of interplatform concordance in terms of genes identified as being differentially expressed [31]. Several of the recognized confounding influences on peripheral blood gene expression were controlled for in the present study [30], including age (by matching in the case-control series), medication use (generally none), and time of day at which blood sampling was conducted (standardized). In addition, we previously reported no significant differences in leukocyte subpopulations between these subject groups [11].

In the DIOS cohort, we have already established that PIFS is a stereotyped illness complex, consistent with the diagnostic criteria for CFS, with a case rate of 11% of subjects at 6 months after the onset of infection [5]. The prospective, population-based research design in the present study can be contrasted with traditional CFS research, which has focused almost exclusively on cross-sectional studies of subjects recruited from tertiary referral clinics. Such subjects feature clinical heterogeneity and chronicity, which are likely to reflect diversity in risk factors, illness course, and pathophysiology [9, 32, 33]. This heterogeneity is likely to be a major reason why the pathogenesis of CFS remains largely unknown, despite several decades of hypothesis-driven research [13, 14]. The PIFS model used here therefore provides a unique opportunity to critically examine the popular hypotheses on the pathogenesis of CFS.

The findings of the present study provide preliminary evidence for the potential of studies of peripheral blood gene expression to identify biomarkers for the major symptoms of the PIFS illness and to open new investigative pathways for studies of pathogenesis. The gene expression signature identified by cluster analysis on the baseline samples generally predicted subsequent PIFS status. However, this signature should be regarded as exploratory only, because it was not unique to the subjects who went on to a PIFS illness. In addition, a cross-sectional analysis of the gene expression data set at 6–9 months after the onset of infection could not reliably distinguish subjects with PIFS from those who had recovered uneventfully from IM.

The genes of interest associated with the major symptoms of PIFS included several with functional roles in metal ion binding within the cell (*CBFA2T2, NDUFS2, CHEK2, ZNF596, MT1X, ZBTB41,* and *ZDHHC3*); immune response pathways (*PRDX1, SCARB1, SH2D1B, RAE1, ASAH2,* and *IL11RA*); hormonal responses (*IGF2AS* and *ACBD3*); and neuronal pathways (*SORCS3* and *KCND3*). The association between each of these genes and PIFS was validated by demonstrating first that the expression in subjects who had recovered from IM differed from those with ongoing symptoms and second that the pattern of change in gene expression over time was associated in a consistent fashion with the longitudinal course of illness (either with ongoing PIFS or with recovery).

It may be noteworthy that none of these candidate genes are shared with those identified in previous studies of patients with CFS, although similar biological processes have been implicated [34–38]. However, these previous studies were cross-sectional and were based on comparison of patients with long-standing CFS, with the likely heterogeneity inherent in that patient group [9, 32]. For instance, a recent analysis of the clinical phenotypes within the diagnosis of CFS in a population-based sample and the associated peripheral blood gene expression pattern revealed at least 5 patient subgroups, each associated with relatively distinct gene expression signatures [39]. Nevertheless, the common biological process identified in these various studies is the transport of iron, zinc, and copper ions, and, in terms of functional pathway, it appears that immune response genes and neuronal genes are commonly expressed. Because these processes and pathways constitute a large proportion of all well-characterized genes in the transcriptome, the fundamental premise of this study—that peripheral blood gene expression will inform a better understanding of the pathogenesis of CFS—remains speculative.

Additional factors beyond alterations in the pattern of peripheral blood gene expression reflecting the host response to the initial infection are likely to contribute to the duration of illness after IM. These may include behavioral response patterns such as alterations in sleep, exercise, and mood as well as the modulating influences of sex [40], which in turn may also influence peripheral blood gene expression [37, 41]. There were no genes identified here that were shared with those previously identified as exercise induced, either in control subjects or in patients with CFS [37].

Further studies of the genes of interest identified here in an expanded case-control series of subjects followed from the onset of acute IM are therefore needed to verify the association with the illness complex of PIFS and to investigate the impact of behavioral changes on gene expression. Confirmation of these gene expression correlates by real-time polymerase chain reaction in the subjects reported here and in subjects who developed PIFS after having other infections included in the DIOS cohort—as well as in independent postinfective cohorts—may make possible novel investigative approaches to elucidate the pathophysiology of PIFS and CFS.

Acknowledgments

The support of the general practitioners and the diagnostic pathology services in the Dubbo region and the enduring cooperation of the subjects who participated in this research are gratefully acknowledged.

References

1. Callan M, Steven N, Krausa P, et al. Large clonal expansions of CD8 T cells in acute infectious mononucleosis. Nat Med **1996**; 2:906–11.
2. Kieff E, Rickinson A. EBV and its replication. In: Knipe D, Howley P,

eds. Fields virology. Vol. 2. 4th ed. Philadelphia: Lippincott Wilkins & Williams, **2001**:2511–74.
3. White P, Thomas J, Kangro H, et al. Predictions and associations of fatigue syndromes and mood disorders that occur after infectious mononucleosis. Lancet **2001**; 358:1946–54.
4. Buchwald D, Rea T, Katon W, Russo J, Ashley R. Acute infectious mononucleosis: characteristics of patients who report failure to recover. Am J Med **2000**; 109:531–7.
5. Hickie I, Davenport TA, Wakefield D, et al. Viral and non-viral pathogens precipitate post-infective and chronic fatigue syndromes: a prospective cohort study. BMJ **2006**; 333:575–8.
6. Sullivan J, Woda B. X-linked lymphoproliferative syndrome. Immunodefic Rev **1989**; 1:325–47.
7. Aoukaty A, Lee I-F, Wu J, Tan R. CAEBV infection associated with low expression of lair-1 on NK cells. J Clin Immunol **2003**; 23:141–5.
8. Kimura H, Hoshino Y, Kanegane H, et al. Clinical and virological characteristics of chronic active EBV infection. Blood **2001**; 98:280–6.
9. 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.
10. Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A. The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. Ann Intern Med **1994**; 121:953–9.
11. 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.
12. Wakefield D, Lloyd A. Pathophysiology of myalgic encephalitis. Lancet **1987**; 2:918–9.
13. Prins JB, van der Meer JWM, Bleijenberg G. Chronic fatigue syndrome. Lancet **2006**; 367:346–55.
14. Afari N, Buchwald D. Chronic fatigue syndrome: a review. Am J Psychiatry **2003**; 160:221–36.
15. de Lange F, Kalkman J, Bleijenberg G, et al. Neural correlates of the chronic fatigue syndrome—an fMRI study. Brain **2004**; 127:1948–57.
16. Georgiades E, Behan W, Kilduff L, et al. Chronic fatigue syndrome: new evidence for a central fatigue disorder. Clin Sci **2003**; 105:213–8.
17. Segman RH, Shefi N, Goltser-Dubner T, Friedman N, Kaminski N, Shalev AY. Peripheral blood mononuclear cell gene expression profiles identify emergent post-traumatic stress disorder among trauma survivors. Mol Psychiatry **2005**; 10:500–13.
18. Tsuang M, Nossova N, Yager T, et al. Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. Am J Med Genet B Neuropsychiatr Genet **2005**; 133:1–5.
19. Vernon S, Nicholson A, Rajeevan M, et al. Correlation of psych-neuroendocrine-immune (PNI) gene expression with symptoms of acute infectious mononucleosis. Brain Res **2006**; 1068:1–6.
20. 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.
21. Hadzi-Pavlovic D, Hickie I, Hooker A, Ricci C. The IFI: some neurasthenia related scales. Sydney: Academic Department of Psychiatry, St. George Hospital, **1997**.
22. 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.
23. 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.
24. 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.
25. von Korff M, Ustun T, Ormel J, Kaplan I, Simon G. Self-report disability in an international primary care study of psychological illness. J Clin Epidemiol **1996**; 49:297–303.
26. Ojaniemi H, Evengard B, Lee D, Unger E, Vernon S. Impact of RNA extraction from limited samples on microarray results. Biotechniques **2003**; 35:968–73.
27. Whistler T, Unger E, Nisenbaum R, Vernon S. Integration of gene expression, clinical, and epidemiologic data to characterize chronic fatigue syndrome. J Transl Med **2003**; 1:10.
28. Garrett-Mayer E. Overview of standard clustering approaches for gene microarray data analysis. In: Allison D, Page G, Beasley T, Edwards J, eds. DNA microarrays and related genomics techniques. Boca Raton, FL: Chapman & Hall CRC, **2006**:131–58.
29. Zhang B, Kirov S, Snoddy J. Webgestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res **2005**; 33:W741–8.
30. Whitney A, Diehn M, Popper SJ, et al. Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci USA **2003**; 100:1896–901.
31. MAQC Consortium. The Microarray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol **2006**; 24:1151–61.
32. 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.
33. Lloyd AR. Chronic fatigue and chronic fatigue syndrome: shifting boundaries and attributions. Am J Med **1998**; 105:7S–10S.
34. 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.
35. 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.
36. Steinau M, Unger ER, Vernon SD, Jones JF, Rajeevan MS. Differential-display PCR of peripheral blood for biomarker discovery in chronic fatigue syndrome. J Mol Med **2004**; 82:750–5.
37. Whistler T, Jones JF, Unger ER, Vernon SD. Exercise responsive genes measured in peripheral blood of women with chronic fatigue syndrome and matched control subjects. BMC Physiol **2005**; 5:5.
38. 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.
39. 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.
40. White PD. What causes chronic fatigue syndrome? BMJ **2004**; 329: 928–9.
41. Rossi EL. Psychosocial genomics: gene expression, neurogenesis, and human experience in mind-body medicine. Adv Mind Body Med **2002**; 18:22–30.