

Utility of the blood for gene expression profiling and biomarker discovery in chronic fatigue syndrome

Suzanne D. Vernon*, Elizabeth R. Unger, Irina M. Dimulescu, Mangalathu Rajeevan and William C. Reeves

Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA

Abstract. Chronic fatigue syndrome (CFS) is a debilitating illness lacking consistent anatomic lesions and eluding conventional laboratory diagnosis. Demonstration of the utility of the blood for gene expression profiling and biomarker discovery would have implications into the pathophysiology of CFS. The objective of this study was to determine if gene expression profiles of peripheral blood mononuclear cells (PBMCs) could distinguish between subjects with CFS and healthy controls. Total RNA from PBMCs of five CFS cases and seventeen controls was labeled and hybridized to 1764 genes on filter arrays. Gene intensity values were analyzed by various classification algorithms and nonparametric statistical methods. The classification algorithms grouped the majority of the CFS cases together, and distinguished them from the healthy controls. Eight genes were differentially expressed in both an age-matched case-control analysis and when comparing all CFS cases to all controls. Several of the differentially expressed genes are associated with immunologic functions (e.g., CMRF35 antigen, IL-8, HD protein) and implicate immune dysfunction in the pathophysiology of CFS. These results successfully demonstrate the utility of the blood for gene expression profiling to distinguish subjects with CFS from healthy controls and for identifying genes that could serve as CFS biomarkers.

Keywords: Gene expression, CFS, cDNA array analysis, peripheral blood

1. Introduction

Chronic fatigue syndrome (CFS) is an illness characterized by debilitating fatigue, impaired concentration and memory, sleep disturbances, and pain [8] and it affects approximately 500 per 100,000 adults in the United States [15]. CFS presents a unique challenge for health care providers, public health officials, and patients because the diagnosis is based on self-reported symptoms and requires exclusion of medical or psychiatric diseases that could potentially explain the illness. Once all other medically explainable illnesses have

been excluded, an individual with CFS is diagnosed by having 6 months or greater persistent or relapsing fatigue and at least four concurrent symptoms including: impaired memory, sore throat, tender lymph nodes, muscle pain, joint pain, new headaches, unrefreshing sleep, and post-exertional malaise [8]. Some features of CFS resemble diseases associated with chronic infection [3,6,13,17,23], immunologic perturbation [16, 18,22], and neuroendocrine disorders [12] but to date, no one etiology or pathology has been defined.

One reason that CFS remains an enigma to both medical and research communities is lack of a known or accessible anatomic lesion. We hypothesized that peripheral blood would serve as a representative sample of the systemic state allowing for evaluation and profiling of multiple pathologic and physiologic pathways. Peripheral blood is an easy and noninvasive sample to collect. Demonstration of the utility of the blood for pro-

* Address for correspondence: Suzanne D. Vernon, Ph.D., Centers for Disease Control and Prevention, 1600 Clifton Road Mail, Stop A-15, Atlanta, 30333, USA. Tel.: +1 404 639 1338; Fax: +1 404 639 2779; E-mail: svernon@cdc.gov.

filing and biomarker discovery in CFS has implications for other medically unexplained diseases. Our purpose was to demonstrate that peripheral blood could serve as a sample for gene expression profiling. We selected a small but well-characterized group of CFS cases and controls and examined the expression of 1764 genes. The approach successfully distinguished the majority of CFS cases from controls and demonstrated the utility of peripheral blood for identifying biomarkers as well as shedding light on etiologic pathways.

2. Materials and methods

2.1. Study subjects

Subjects and study design have been presented in detail elsewhere [18,20]. In brief, the original study was a case/control design that included subjects with CFS defined by the 1988 case definition [14] and age, race, and sex matched controls selected randomly from the Atlanta, Georgia metropolitan area. Five CFS cases and seventeen controls had remaining cryo-preserved PBMC aliquots available for gene expression analysis with cDNA arrays; all these subjects were white women age 24 through 51. The five CFS cases had been ill from 2.5 to 6.5 years. Human experimentation guidelines of the U.S. Department of Health and Human Services were followed in the conduct of this study. All study participants were volunteers who gave informed consent.

2.2. Specimens

Specimen collection has also been described [18]. Briefly, cases and controls had blood samples collected before 10 a.m. in vacutainers containing citric acid. PBMCs were isolated on lymphocyte separation medium (Organon Teknika, Durham, NC), dispensed into 5 million cell aliquots and stored in liquid nitrogen under conditions to maintain viability.

2.3. RNA preparations

Total RNA was extracted from the cryo-preserved PBMC aliquots from the remaining five CFS cases and seventeen controls by the modified guanidinium thiocyanate method [4]. On average, 5 – 10 μg of total RNA was obtained from each 5 million PBMC aliquot. Residual DNA was removed with DNase digestion, 1 U DNase I (GenHunter Corp., Nashville, TN) per 10 μg

total RNA for 15 minute at room temperature, followed by phenol-chloroform extraction and ethanol precipitation. We usually recovered 80% of the original total RNA preparation. All RNA samples were examined for RNA integrity and absence of DNA by denaturing agarose gel electrophoresis, and were quantified by UV-spectrophotometry.

2.4. Probe synthesis

Digoxigenin labeled double-stranded cDNA probes were synthesized using the SMART PCR cDNA Synthesis Kit and the Advantage cDNA PCR Kit (CLONTECH, Palo Alto, CA) as previously described [25]. One μg total RNA for each sample was also included in the labeling reaction to monitor for any DNA carry over after extraction (no reverse transcriptase control). One μl of the 100 μl PCR product from each sample and corresponding DNA carry over control were evaluated for dig-11-dUTP incorporation by agarose gel electrophoresis. Fifty μl of the labeled PCR product was used as the probe in the hybridization.

2.5. Hybridization and chemiluminescent detection

Samples were hybridized to the Atlas Human cDNA Expression Array with 588 genes (CLONTECH) and the Atlas Human 1.2 Array II with 1176 genes (CLONTECH) under previously described conditions [25].

2.6. Analysis

The chemiluminescent signal was detected with a one-hour exposure to film. Films were scanned on a flatbed scanner and Tagged Image File Formats were generated. Image files were loaded into BioNumerics (Applied Maths, Kortrijk, Belgium) for signal intensity quantification. Due to the variability in background and noise level between films, intensity values for each image were normalized based on the negative and positive controls on each filter. Setting the lowest intensity blank control on each filter to zero, and the highest intensity positive control to one hundred, the BioNumerics program proportionally scaled the intensities between 0 and 100. The program subsequently quantified all signal intensities between 0 and 100. The threshold value for positive signals for each image was determined as the average of the three lowest intensity negative controls plus five standard deviations. Intensity values below this threshold were set to 0.01. The

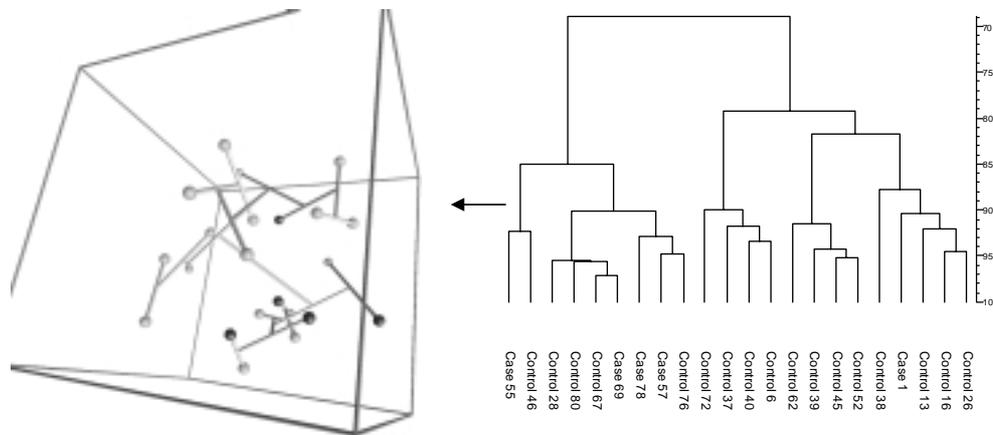


Fig. 1. Cluster analysis of gene expression data from the five CFS cases and the seventeen controls. The bar at the right of the dendrogram indicates the percent similarity. The CFS cases are identified as the red squares and the controls are the green squares in the hierarchical tree. This red green color scheme is maintained in the three dimensional plot of the multidimensional scaling analysis.

positive normalized intensity values were transformed by \log_2 for analysis.

Three hundred-fifty one genes were negative in 18 (80%) of the 22 subjects and were excluded from further analysis. Subjects were grouped based on disease status (CFS versus healthy) and further stratified by age. The overall agreement for each comparison was evaluated using the Pearson correlation coefficient. The criterion for differential gene expression was greater than four-fold variation in gene intensity. The variation in the expression of each gene was further examined with the nonparametric Wilcoxon test. All data analysis including cluster analysis and multidimensional scaling was performed using BioNumerics (Applied Maths), Microsoft Excel (Microsoft Office XP, Microsoft Corporation), and Statistica (StatSoft, Tulsa, OK).

3. Results

We used hierarchical clustering and multidimensional scaling to determine if the PBMC expression profiles of the CFS cases were distinguishable from healthy controls. Unsupervised cluster analysis yielded two major branches in the dendrogram (Fig. 1). One branch contained four (80%) of five CFS cases and five healthy controls. The other branch contained twelve (71%) of seventeen healthy controls and one CFS case. Multi-dimensional scaling of the intensity values illustrated that the majority of CFS cases (red spheres) grouped together and were distant and distinct from most healthy controls (green spheres) (Fig. 1). The five controls that clustered with the four CFS cases in the bottom branch

were the age-matched controls for three of these cases (shown with identifying numbers 55, 69 and 78). Since all samples were from white females, age may be a significant variable in clustering the profiles. However, not all age-matched controls clustered with their corresponding case [e.g., CFS case 57 (27 years old) did not cluster with her matched control 26 (25 years old)].

The intensity values of each of the genes from the CFS group were compared to the corresponding gene in the control group using the Wilcoxon rank sum test. Nineteen genes were identified as different between these two groups (Table 1). The mean intensity value for each significant gene for the CFS group and the control group are shown to give a sense of the difference. To control for age, we repeated the analysis using all CFS cases and two age-similar controls for each case. In the age-matched analysis, nine genes were different between the CFS group and the control group (Table 1). Seven genes were common to both analyses (highlighted in grey). Both comparisons had genes that were unique to each group.

Discriminant analysis was also used to compare gene expression of the CFS cases to the healthy control group to detect the most discriminating variables (genes). Four genes were most discriminating in both the total and age-matched analysis: CDK-interacting protein, transcription factor ETR101, CMRF35 and ICAM2 (Fig. 2). The comparison of CFS group to all seventeen controls detected four additional discriminating genes (Fig. 2a) while the comparison of the five CFS cases to the ten age-matched controls detected seven discriminating genes (Fig. 2b). Both discriminant and non-parametric methods detected some of the same genes

Table 1

Comparison of all CFS cases to all 17 healthy controls (top) or to 10 age-matched healthy controls (bottom). The genes highlighted in bold are common to both total and age-matched comparisons

Gene name	Genbank #	Case mean	Control mean	p value*
<i>All CFS cases and all controls</i>				
CMRF35 antigen precursor	X66171	29	8	0.004
integrin alpha 6	X53586; X59512	2	9	0.004
zinc finger protein 145 (ZNF145)	Z19002; S60093	7	1	0.008
Huntington's disease protein (HD protein)	L12392	8	1	0.01
interleukin-8 precursor (IL-8)	Y00787	5	20	0.01
membrane-associated phospholipase A2 precursor	M22430; J04704	10	1	0.01
dioxin-inducible cytochrome P450 (CYP1B1)	U03688	2	18	0.02
ATP-regulated potassium channel ROM-K	U12541	11	4	0.008
NF-AT4c	L41067	9	2	0.008
alpha2A-Adrenoceptor	M18415	9	3	0.02
Grave's disease carrier protein (GDC)	M31659	11	4	0.02
DNA mismatch repair protein hmlh1	U07418	67	50	0.03
dynactin 150-kDa isoform	X98801	14	6	0.03
transcription factor ETR101	M62831	63	38	0.03
ADP-ribosylation factor 1	M36340	58	36	0.04
ataxin-1; spinocerebellar ataxia type 1 protein	X79204	13	6	0.04
lymphocyte cytosolic protein 1 (LCP-1)	M22300	98	86	0.04
fibronectin receptor beta subunit	X07979	5	23	0.05
intercellular adhesion molecule 2 precursor (ICAM2)	X15606	53	30	0.05
<i>All CFS cases and age matched controls</i>				
Huntington's disease protein (HD protein)	L12392	8	0	0.01
zinc finger protein 145 (ZNF145)	Z19002; S60093	7	1	0.02
CMRF35 antigen precursor	X66171	29	12	0.03
integrin alpha 6	X53586; X59512	2	7	0.03
membrane-associated phospholipase A2 precursor	M22430; J04704	10	2	0.03
dioxin-inducible cytochrome P450 (CYP1B1)	U03688	2	11	0.04
interleukin-8 precursor (IL-8)	Y00787	5	18	0.04
beta-adrenergic receptor kinase 2 (beta-ARK2)	X69117	12	26	0.02
protein kinase C eta type (NPKC-eta)	M55284	9	3	0.04

*Wilcoxon exact p value.

as differentially expressed between the CFS group and the control group. It is notable that the CMRF35 antigen precursor gene was detected by all methods in both the age-matched and non age-matched comparison.

4. Discussion

We measured gene expression of PBMCs from CFS cases and controls. We chose to profile the blood because it is a sample that reflects many ongoing systemic pathophysiologic processes. Cluster analysis of the PBMC expression profiles separated the majority of CFS cases from the majority of controls. Similar gene expression profiling studies have been used to identify prognostic biomarkers in a variety of malignancies (e.g., lymphoma, prostate, breast and cervix) [1,5,21]. Each of these studies measured the gene expression of tissues taken directly from lesions and compared profiles to those of normal tissue. Our study demonstrates the utility of the blood for gene expression profiling on an illness without a known lesion. A similar ap-

proach may be useful in studying diseases with lesions that are difficult to sample as well as other unexplained illnesses that lack anatomic lesions.

Both cluster analysis and multidimensional scaling grouped and separated CFS from controls. The method of cluster analysis we employed uses a joining or tree clustering algorithm that forms branches based on similarity or distance. In Fig. 1, five controls clustered with the CFS cases. These included the age-matched controls for three of the CFS cases. This indicated that in addition to disease status, age could affect the expression of certain genes and consequently influence how individuals cluster. We controlled for other variables that could impact PBMC gene expression such as sex, race and time of sample collection. It is important to consider and attempt to control for as many variables as possible when using classification algorithms such as clustering. It is also important to validate results using more than one classification algorithm.

All five CFS cases were white women who had been ill for many years. The reason that the one CFS case did not cluster with the other cases cannot be deter-

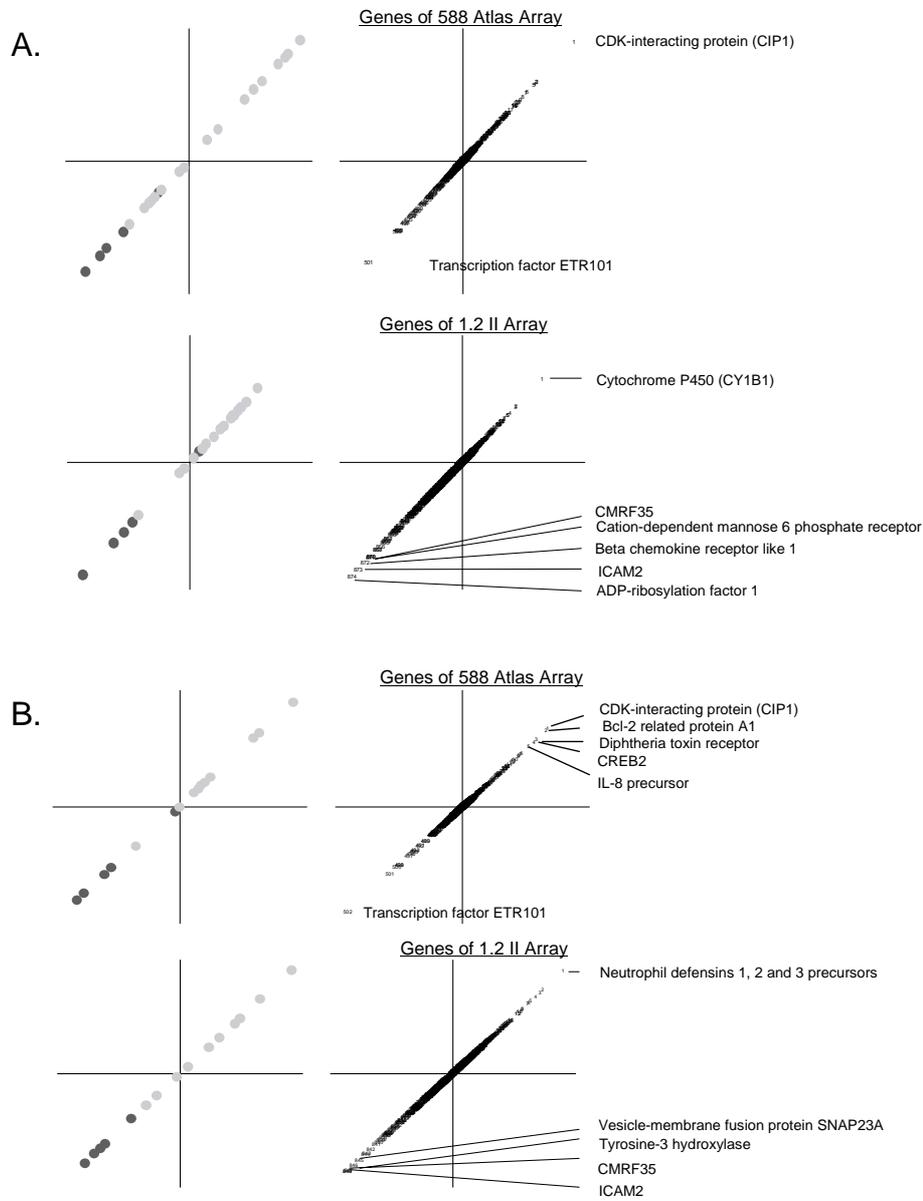


Fig. 2. Discriminant analysis on genes from the 588 Atlas and the 1.2 II array. Figure 2A illustrates discriminant analysis of the five CFS cases in one group (red spheres) and the 17 controls in the other group (green spheres). The top set of graphs show the analysis on the 514 positive genes from the 588 Atlas array and the bottom set of graphs are the analysis on the 899 positive genes from the 1.2 II array. The most discriminating genes are indicated. Figure 2B illustrates discriminant analysis of the five CFS cases in one group (red spheres) and the 10 age-similar controls in the other group (green spheres). The top set of graphs show the analysis on the 514 positive genes from the 588 Atlas array and the bottom set of graphs are the analysis on the 899 positive genes from the 1.2 II array. Again, the most discriminating genes are indicated.

mined. This could point to heterogeneity of CFS and the occurrence of CFS subgroups (i.e., disease stratification) [24]. Recognizing that CFS may not be one entity, the current case definition encourages investigators to stratify cases based on mode of onset, as well as other potential distinguishing characteristics [8].

Differentially expressed genes were identified be-

tween CFS cases versus all controls or age-matched controls using the Wilcoxon test and discriminant analysis. The non-parametric analysis identified seven genes that distinguished CFS case group from both age-matched and unmatched control groups. Of these seven genes, four (CMRF35, HD protein, phospholipase A2 and ZNF145) were more highly expressed in

CFS cases than in controls. It is noteworthy that one gene, the CMRF35 antigen precursor, was detected as differentially expressed by all analytical approaches. This gene encodes a cell membrane antigen that is a member of the immunoglobulin superfamily. The CMRF35 gene is expressed as a receptor or ligand to varying degrees in subsets of immune cells (T, B, natural killer, and myeloid cells) [7]. As an immunoglobulin superfamily antigen, the CMRF35 precursor plays a role in regulating lytic and cytokine expression capabilities of immune cells and is thought to control interactions between T cells and antigen presenting cells or target (virus-infected or mutated) cells that have to be killed [10]. Interestingly, immune function studies from this same group of CFS case and control samples has shown a significant difference in natural killer cell surface markers with CFS cases having less expression of CD2 than controls [18].

None of the genes identified in this analysis have been previously characterized in CFS cases. These will form the basis for developing and testing novel hypothesis of CFS pathogenesis. In addition to the CMRF35 antigen, several of the differentially expressed genes detected here suggest an immunological basis for CFS. Interleukin-8 is a potent proinflammatory chemotactic factor and is down-regulated in the CFS group compared to controls. The CMRF35 antigen was highly expressed in the CFS group and as discussed above is thought to regulate (inhibit) cytokine expression [2]. Integrins are essential for adhesion and function in cell migration, cell proliferation and differentiation. This gene was not as highly expressed in the CFS group compared to the control group. The Huntingtin protein may play a role in apoptosis. It has recently been shown that the mutant Huntingtin protein up-regulates expression of the cell death gene caspase-1 [9]. All of these genes implicate immune dysfunction in the pathophysiology of CFS.

Simultaneous evaluation of thousands of expressed genes in the peripheral blood represents a powerful approach for characterizing illnesses without a known or accessible lesion. This study is a successful "proof-of-concept" for this approach. However, the small numbers of samples that were available for analysis and the fact that we only assessed the expression of 1,764 genes must temper the strength of our conclusions about the significance of any of the differentially expressed gene detected here and its role in the pathophysiology of CFS. In addition, the differential expression of the genes identified could not be independently validated in these samples because there was no residual RNA avail-

able [19]. Further studies using peripheral blood samples from larger numbers of well-characterized subjects are necessary to minimize individual differences and experimental variability. In addition, glass microarrays provide the possibility of studying a broader spectrum of genes. These studies are in progress and will potentially yield a clearer profile of aberrant pathophysiological pathways and identify diagnostic biomarkers of CFS.

Acknowledgments

We would like to thank Dr. Rosane Nisenbaum for thoughtful discussion of the data analysis and results.

References

- [1] A.A. Alizadeh, M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald et al., Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, *Nature* **403** (2000), 503–511.
- [2] F. Borrego, J. Kabat, D.K. Kim, L. Lieto, K. Maasho, J. Pena, R. Solana and J.E. Coligan, Structure and function of major histocompatibility complex (MHC) class I specific receptors expressed on human natural killer (NK) cells, *Mol Immunol* **38** (2002), 637–660.
- [3] D. Buchwald, R.L. Ashley, T. Pearlman, P. Kith and A.L. Komaroff, Viral serologies in patients with chronic fatigue and chronic fatigue syndrome, *J Med Virol* **50** (1996), 25–30.
- [4] P. Chomczynski and N. Sacchi, Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform, *Anal Biochem* **162** (1987), 156–159.
- [5] S.M. Dhanasekaran, T.R. Barrette, D. Ghosh, R. Shah, S. Varambally K. Kurachi et al., Delineation of prognostic biomarkers in prostate cancer, *Nature* **412** (2001), 822–826.
- [6] B. Evengard, T. Briese, G. Lindh, S. Lee and W.I. Lipkin, Absence of evidence of Borna disease virus infection in Swedish patients with chronic fatigue syndrome, *J Neurovirol* **5** (1999), 495–499.
- [7] C. Fraser, D. Howie, M. Morra, Y. Qiu, C. Murphy, Q. Shen et al., Identification and characterization of SF2000 and SF2001, two new members of the immune receptor SLAM/CD2 family, *Immunogenetics* **53** (2002), 843–850.
- [8] K. Fukuda, S.E. Straus, I. Hickie, M.C. Sharpe, J.G. Dobbins and A.L. Komaroff, the International Chronic Fatigue Syndrome Group, The chronic fatigue syndrome: a comprehensive approach to its definition and study, *Ann Intern Med* **121** (1994), 953–959.
- [9] F.G. Gervais, R. Singaraja, S. Xanthoudakis, C.A. Gutekunst, B.R. Leavitt, M. Metzler et al., Recruitment and activation of caspase-8 by the Huntingtin-interacting protein Hip-1 and a novel partner Hip1, *Nat Cell Biol* **4** (2002), 95–105.
- [10] C.E. Grossi, E. Ciccone, C. Tacchetti, G. Santoro and G. Anastasi, Anatomy of the immune system: facts and problems, *Ital J Anat Embryol* **105** (2000), 97–124.
- [11] I. Hedenfalk, D. Duggan, Y. Chen, M. Radmacher, M. Bittner, R. Simon et al., Gene expression profiles in hereditary breast cancer, *N Engl J Med* **344** (2001), 539–548.

- [12] C. Heim, U. Ehlert and D.H. Hellhammer, The potential role of hypocortisolism in the pathophysiology of stress-related bodily disorders, *Psychoneuroendocrinology* **2** (2000), 1–35.
- [13] W. Heneine, T.C. Woods, S.D. Sinha, A.S. Khan, L.E. Chapman, L.B. Schonberger and T.M. Folks, Lack of evidence for infection with known human and animal retroviruses in patients with chronic fatigue syndrome, *Clin Infect Dis* **18S** (1994), 121–125.
- [14] G.P. Holmes, J.E. Kaplan, N.M. Gantz, A.L. Komaroff, L.B. Schonberger, S.E. Straus et al., Chronic fatigue syndrome: A working case definition, *Ann Intern Med* **108** (1988), 387–389.
- [15] L.A. Jason, J.A. Richman, A.W. Rademaker, K.M. Jordan, A.V. Plioplys, R.B. Taylor et al., A community-based study of chronic fatigue syndrome, *Arch Intern Med* **159** (1999), 2129–2137.
- [16] N.G. Klimas, F.R. Salvato, R. Morgan and M.A. Fletcher, Immunologic abnormalities in chronic fatigue syndrome, *J Clin Microbiol* **28** (1992), 1404–1410.
- [17] A.C. Mawle, R. Nisenbaum, J.G. Dobbins, H.E. Gary, J.A. Stewart, M. Reyes et al., Seroepidemiology of chronic fatigue syndrome: A case-control study, *Clin Infect Dis* **21** (1995), 1386–1389.
- [18] A.C. Mawle, R. Nisenbaum, J.G. Dobbins, H.E. Gary, J.A. Stewart, M. Reyes et al., Immune responses associated with chronic fatigue syndrome: A case-control study, *J Infect Dis* **175** (1997), 136–141.
- [19] M.S. Rajeevan, S.D. Vernon, N. Taysavang and E.R. Unger, Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR, *J Mol Diagn* **3** (2001), 26–31.
- [20] M. Reyes, J.G. Dobbins, A.C. Mawle, L. Steele, H.E. Gary, H. Malani et al., Risk factors for chronic fatigue syndrome: a case control study, *J Chronic Fatigue Syndrome* **2** (1996), 17–33.
- [21] C. Shim, W. Zhang, C.H. Rhee and J.H. Lee, Profiling of differentially expressed genes in human primary cervical cancer by complementary DNA expression array, *Clin Cancer Res* **4** (1998), 3045–3050.
- [22] S.E. Straus, S. Fritz, J.K. Dale, B. Gould and W. Strober, Lymphocyte phenotype and function in the chronic fatigue syndrome, *J Clin Immunol* **13** (1993), 30–40.
- [23] C.W. Swanink, W.J. Melchers, J.W. van der Meer, J.H. Ver-coulen, G. Bleijenberg, J.F. Fennis and J.M. Galama, Enteroviruses and the chronic fatigue syndrome, *Clin Infect Dis* **19** (1994), 860–864.
- [24] R.R. Taylor, F. Friedberg and L.A. Jason, *A clinician's guide to controversial illnesses: Chronic fatigue syndrome, Fibromyalgia, and Multiple Chemical Sensitivities*, Sarasota, FL: Professional Resource Press 2001.
- [25] S.D. Vernon, E.R. Unger, M. Rajeevan, I.M. Dimulescu, R. Nisenbaum and C. Campbell, Reproducibility of alternative probe synthesis approaches for gene expression profiling with arrays, *J Molecular Diagnostics* **2** (2000), 124–127.



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

